

Skills in the Labor Market: Measurement, Formation, and Returns



Florian Schoner
Department of Economics
LMU Munich

A thesis submitted for the degree of
Doctor oeconomiae publicae

2026

Skills in the Labor Market: Measurement, Formation, and Returns

Inaugural-Dissertation

zur Erlangung des akademischen Grades
Doctor oeconomiae publicae (Dr. oec. publ.)

an der Volkswirtschaftlichen Fakultät

an der Ludwig-Maximilians-Universität München

vorgelegt von

Florian Schoner

2026

Referent: Prof. Dr. Ludger Woessmann

Korreferent: Prof. Dr. Fabian Kosse

Promotionsabschlussberatung: 28. Januar 2026

“Confusion, I was taught, is the beginning of understanding.”

— Yongey Mingyur Rinpoche

Acknowledgements

This thesis is the product of a journey that offered intellectual rewards and personal joy, but also trials and setbacks. I owe heartfelt thanks to all who made it special and stood by me along the way.

First, I want to express my gratitude to my advisors, Ludger Woessmann and Fabian Kosse. Ludger, I learned immensely from our collaborations. Thank you for supporting my research visit abroad, for the chance to contribute to policy projects I care about, and for creating such an excellent environment for research. Fabian, your supervision of my master's thesis was pivotal in my decision to come to Munich. I learned a great deal from you about the "hidden curriculum" of academia, and I am grateful for your steady support and encouragement.

I was fortunate to call ifo—and BI in particular—my professional home for the past five years. Thank you to all current and former colleagues for creating such a warm and supportive environment. Special thanks to Sarah Leyh, Katharina Wedel, Raphael Brade, and Moritz Seebacher for their support and friendship. Sarah, Simon ter Meulen, and Giuseppe di Giacomo were also excellent cycling buddies. I am grateful to Caterina Pavese for her open door and encouragement, and to Julius Berger for companionship (and for keeping the coffee supply stable). Finally, heartfelt thanks to Franziska Binder and Ulrike Baldi-Cohrs for their constant support behind the scenes and for making this place so welcoming.

I am grateful for the opportunity to visit UC Berkeley with financial support from ifo and CESifo. I am indebted to Enrico Moretti for hosting me. The visit not only advanced my development as a researcher but also led to new friendships—thank you, Giuseppe.

This thesis would not have been possible without my brilliant co-authors. I learned an enormous amount from Lukas Mergele and Larissa Zierow in the early phase. Without your steady encouragement and personal support, I would not have made it through the first year. With David Dorn, Moritz Seebacher, and Lisa Simon, I had the privilege of working in an outstanding team. I thank David for his steadfast commitment and Moritz for his dependability. I also thank Philipp Lergetporer for his trust, and Hanna Brosch for her steady dedication and pragmatism.

I am deeply grateful to the family and friends who accompanied me on this journey and stood by me in difficult times. I thank my parents, especially my

mother, Birgit, whose steady encouragement helped me climb each rung of the education ladder. I'm grateful to Steffen for his encouragement and for making his home a standing invitation. In Munich, I was lucky to share flats with Mara, Daniel, Jonas, Nico, and Neele—thank you for making it feel like home. I am also grateful for close friends nearby: Julia and Stefan, thank you for sticking with me for more than ten years. Jonas and Maike, you gave me a place to recharge, to feel seen, and to laugh—thank you. Nora, Sarah, and Marvin were always within reach, even if not physically.

Contents

List of Figures	xiii
List of Tables	xv
1 General Introduction	1
1.1 Skills and Human Capital in the Labor Market	1
1.2 Data	4
1.3 Identification, Estimation Methods, and Overview	6
2 Multidimensional Skills on LinkedIn Profiles: Measuring Human Capital and the Gender Skill Gap	9
2.1 Introduction	9
2.2 Related Literature	12
2.3 LinkedIn Data and the Measurement of Skills	15
2.3.1 LinkedIn Profiles	16
2.3.2 Cross-Sectional Analysis	16
2.3.3 Self-Reported Skills	17
2.4 Definition of Variables and Sample	20
2.4.1 Skill Clusters	21
2.4.2 Additional Variables Observed in LinkedIn Profiles	26
2.4.3 Variables Inferred from LinkedIn Profiles and External Sources	27
2.4.4 Sample Selection and Representativeness	29
2.5 Human Capital Investments and Skills	30
2.5.1 Skills by Age and Experience	32
2.5.2 Skills by Education: Degrees, Fields of Study, and College Quality	36
2.6 Skills and Job-based Earnings	39
2.6.1 Number of Skills, Skill Composition, and Earnings	39
2.6.2 Multidimensional Skills and Earnings	41
2.6.3 Do Skills Explain More Earnings Variation than Standard Human Capital Measures?	43
2.7 The Gender Skill Gap	48

2.7.1	Skills by Gender	48
2.7.2	Skills and the Gender Earnings Gap	58
2.8	Conclusions	62
2.A	Appendix A: Additional Tables and Figures	65
2.B	Appendix B: Results for LinkedIn Users with Recently Updated Profiles	81
2.C	Appendix C: Results with ACS Population Weights	86
3	Grading Student Behavior	91
3.1	Introduction	91
3.2	Institutional Setting and Theoretical Considerations	95
3.2.1	Institutional Setting	95
3.2.2	Theoretical Considerations	98
3.3	Data	99
3.4	Empirical Strategy	107
3.5	Results	111
3.5.1	Comportment Grading and the School-to-Work Transition	111
3.5.2	Non-Cognitive Skills and Reading Skills	113
3.5.3	Discussion of Main Results	116
3.6	Potential Explanations	117
3.7	Costs of Comportment Grading	123
3.8	Conclusion	124
3.A	Appendix A: Policy Background	127
3.B	Appendix B: Treatment Effect Estimands	131
3.C	Appendix C: Teacher Survey	133
3.D	Appendix D: Robustness Checks	134
3.D.1	School-to-Work Transition	134
3.D.2	Non-cognitive Skills and Reading Skills	138
3.D.3	Effect on Composition of the Academic Track	140
3.E	Appendix E: Federal State of Schooling	141
4	Worker Beliefs About Firm Training	143
4.1	Introduction	143
4.2	Conceptual Considerations and Institutional Background	148
4.2.1	Conceptual Considerations	148
4.2.2	Institutional Background: Firm Training in Germany	151
4.3	The Survey	153
4.3.1	Data Collection	153
4.3.2	Measuring Actual and Intended Firm Training Participation	155
4.3.3	Eliciting Workers' Beliefs	157

4.3.4	Information-Provision Experiment	162
4.4	Results	163
4.4.1	Beliefs about Returns to Firm Training	163
4.4.2	Return Beliefs Predict Intended and Realized Training Participa- tion	168
4.4.3	Belief Gaps Between Lower- and Higher-Skilled Workers . .	170
4.4.4	Experimental Results	175
4.5	Discussion and Conclusion	177
4.A	Appendix A: Additional Figures and Tables	181
4.B	Appendix B: Information Experiment	187
4.C	Appendix C: Robustness to Alternative Definition of Higher-skilled	191
4.D	Appendix D: Robustness to Winsorization	193
4.E	Appendix E: Robustness to Bad Data	194
4.F	Appendix F: Survey	198
4.F.1	Measuring Firm Training Participation	198
4.F.2	Belief Elicitation	200
4.F.3	Information Experiment	201
5	Labor Market Returns to Skills: Evidence from LinkedIn	203
5.1	Introduction	203
5.2	Data and Descriptive Statistics	207
5.2.1	LinkedIn Profiles	208
5.2.2	Panel Creation and Sample	208
5.2.3	Skills	210
5.2.4	Interpreting Skill Additions as Human Capital Accumulation	211
5.2.5	Job Mobility	213
5.2.6	Earnings	214
5.3	Empirical Framework	218
5.3.1	Identification Strategy	218
5.3.2	Assessing Identification	220
5.3.3	Threats to Identification	220
5.3.4	Empirical Strategy	222
5.4	Results	226
5.4.1	Job Mobility	226
5.4.2	Earnings	227
5.5	Robustness	229
5.5.1	Placebo	229
5.5.2	Sensitivity	230
5.5.3	Further Robustness Checks	232

5.6	Summary and Concluding Remarks	234
5.A	Appendix A: Data and Variance Decomposition	236
5.A.1	Further Data	236
5.A.2	Decomposing the contribution of changes to positions to explained variance in earnings trajectories	237
5.B	Appendix B: Robustness Checks	239
5.B.1	Intensive Margin Treatment	239
5.B.2	Staleness	242
5.B.3	Covid	243
5.B.4	Reverse Causality	244
	Overview of Tools Used	247
	References	249

List of Figures

2.1	Age-skill profile: Number of reported skills by age	33
2.2	Number of reported skills by experience and labor-market attachment	34
2.3	Skill composition by age: Shares of general, managerial, and specific skills	35
2.4	Number of reported skills and skill composition by education . . .	37
2.5	Job-based earnings by number of reported skills and skill composition: Unweighted vs. weighted data	40
2.6	Association between job-based earnings and skills	42
2.7	Number of reported skills and skill composition by age and gender.	50
2.8	Alternative age-skill profiles by gender	53
2.9	Skill accumulation by gender, relative female/male work hours, and motherhood	56
A1	Job-based (Revelio Labs) and occupation-based (ACS) earnings measures	74
A2	Comparison of LinkedIn sample to ACS	75
A3	Age-earnings profile for college graduates	76
A4	Skill composition by experience and labor-market attachment . . .	77
A5	Age-earnings profiles for college graduates by gender	78
A6	Skill composition by age and gender conditional on occupation . .	79
A7	Potential experience, actual experience, and labor force attachment by gender	80
B1	Number of reported skills and skill composition by age: Baseline sample vs. recent firm switchers	82
B2	Job-based earnings by number of reported skills and skill composition: Baseline sample vs. recent firm switchers	84
B3	Number of reported skills by age and gender: Baseline sample vs. recent firm switchers	85
C1	Number of reported skills and skill composition by age: Unweighted vs. weighted data	87
C2	Job-based earnings by number of reported skills and skill composition: Unweighted vs. weighted data	89

C3	Number of reported skills by age and gender: Unweighted vs. weighted data	90
3.1	The introduction of comporment grading over time and by state within the sample	105
3.2	Dynamic effect of comporment grading on school-to-work transitions	115
3.3	Teacher survey: Assessment and basis of comporment grading . .	123
A1	Report cards for Jimmy Carter (left) and Lyndon B. Johnson (right)	127
A2	School report card for Harry S. Truman	128
A3	Stylized overview of the German school system	128
A4	By group treatment effect of comporment grading on school-to-work transitions	134
A5	Four period setup: Dynamic effect of comporment grading on school-to-work transitions	135
4.1	Belief elicitation task	159
4.2	Beliefs about non-pecuniary returns	166
4.3	Do return beliefs predict intended firm training participation? . . .	170
4.4	Difference in return beliefs by workers' skill-level	173
4.5	Do return beliefs predict realized firm training participation?	185
4.6	Difference in return beliefs by workers' available skill training . . .	186
4.7	Difference in return beliefs by workers' skill-level (alternative definition)	192
4.8	Beliefs about non-pecuniary returns (only reliable respondents, N = 2962)	195
4.9	Difference in return beliefs by workers' skill-level (only reliable respondents, N = 2962)	196
4.10	Survey Flow	198
5.1	Timeline of sample construction	210
5.2	Average predicted and ACS earnings for treatment and control group over time	217
5.3	Propensity score distributions of trimmed sample	224
5.4	Sensitivity analysis for effect on 2023 predicted earnings	231
5.B.1	Profile Notes: Appendix LinkedIn Profile	246

List of Tables

2.1	Skill clusters: Descriptive statistics	21
2.2	Explaining earnings variation with alternative metrics of human capital	46
2.3	Skills and the gender earnings gap	60
A1	Most frequent skill strings in each skill cluster	65
A2	Sample selection	68
A3	Skill count and composition by experience and education: Regressions	69
A4	Alternative metrics of human capital and job-based earnings	70
A5	Female skill growth as a function of male skill growth and the female/male work hours ratio	74
3.1	Descriptive statistics Mikrozensus	101
3.2	Descriptive statistics SOEP	102
3.3	Descriptive statistics nationwide student assessments	104
3.4	Descriptive statistics NEPS	107
3.5	Effect of compartment grading on school-to-work transitions	113
3.6	Effect of compartment grading on school-to-work transitions using 12 states	114
3.7	Effect of compartment grading on non-cognitive skills and reading skills	116
3.8	Explaining Compartment Grades using Subject Grades	119
3.9	Contemporaneous compartment grading and disciplinary problems as rated by school principal and classroom disruptions as rated by students	120
3.10	Teacher Survey	122
A1	Grading of social and work behavior in selected European countries	129
A2	Teacher guidelines for the evaluation of behavior (excerpt) in the state of Baden-Wuerttemberg	130
A3	Teacher guidelines for the evaluation of behavior in the state of Saxony	130
A4	Descriptive statistics teacher survey: Teacher characteristics	133
A5	Effect of compartment grading on school-to-work transitions: Ro- bustness checks	136
A6	Effect of compartment grading on school-to-work transitions: Ro- bustness checks using treatment assignment in different grades . .	137

A7	Effect of comporment grading on non-cognitive skills and reading skills - without controls	138
A8	Effect of comporment grading on non-cognitive skills and reading skills - Treatment assignment in grade 4	139
A9	Effect of comporment grading on academic track school attendance	140
4.1	Respondent characteristics	155
4.2	Realized and intended firm training participation	157
4.3	Predictives of realized and intended firm training participation . .	158
4.4	Earnings expectations descriptives	165
4.5	Earnings expectations by workers' skill-level	172
4.6	Skill gap in firm training intentions	174
4.7	Information treatment effects on training intentions	176
4.8	Representativeness	181
4.9	Beliefs about (non-)pecuniary returns	182
4.10	Robustness: difference in return beliefs by workers' skill-level . . .	182
4.11	Respondent characteristics separately for higher- and lower-skilled	183
4.12	Firm training participation separately for higher- and lower-skilled	183
4.13	Details on firm training participation separately for higher- and lower-skilled	184
4.14	Skill gap in realized firm training	187
4.15	Balancing	187
4.16	Prior beliefs	188
4.17	Heterogenous treatment effects on training intentions by prior beliefs	188
4.18	Heterogenous treatment effects by workers' age	189
4.19	Heterogenous treatment effects by workers' age (continuous)	190
4.20	Earnings expectations by workers' skill-level (alternative definition)	191
4.21	Skill gap in firm training intentions (alternative definition)	192
4.22	Information treatment effects on training intentions (alternative definition)	193
4.23	Earnings expectations by workers' skill-level (no winsorization) . .	193
4.24	Earnings expectations descriptives (no winsorization)	194
4.25	Earnings expectations descriptives (only reliable respondents, N = 2962)	194
4.26	Earnings expectations by workers' skill-level (only reliable respondents, N = 2962)	195
4.27	Information treatment effects on training intentions (reliable respondents, N = 2962)	197
5.1	Descriptive statistics for across-time variation in skills	212

5.2	Decomposition of predicted earnings change by switch of position characteristic	216
5.3	Balancing table after trimming	225
5.4	OLS estimates for the effect of accumulating skills on job mobility	226
5.5	OLS estimates for the effect of accumulating skills on predicted earnings	228
5.6	OLS estimates for the effect of accumulating skills on ACS earnings	228
5.7	OLS placebo estimates for the effect of accumulating skills on lagged earnings	229
5.B.1	Effect of an Additional Skill on Post-Treatment Earnings	239
5.B.2	Effect of an Additional Skill on Post-Treatment ACS Earnings	239
5.B.3	OLS estimates for the effect of accumulating skills on job mobility (count specification)	239
5.B.4	Earnings and Skill Additions: Two-Part Specification	240
5.B.5	ACS Earnings and Skill Additions: Two-Part Specification	240
5.B.6	OLS estimates for the extensive vs. intensive margin of skill accumulation on job mobility	241
5.B.7	Earnings and Skill Additions: Mutually Exclusive Treatment Bins (ref. = 0)	241
5.B.8	Treatment Effect on Post-Treatment Earnings	242
5.B.9	Treatment effect on Post-Treatment ACS Earnings	242
5.B.10	Treatment Effect on job mobility outcomes	242
5.B.11	Placebo Treatment Effects on Lagged Earnings (LinkedIn vs. ACS)	243
5.B.12	Treatment Effect on Post-Treatment Earnings	243
5.B.13	Treatment effect on Post-Treatment ACS Earnings	244
5.B.14	Treatment Effect on job mobility outcomes	244
5.B.15	Treatment Effect on Post-Treatment Earnings	244
5.B.16	Treatment effect on Post-Treatment ACS Earnings	245
5.B.17	Treatment Effect on job mobility outcomes	245
5.B.18	Placebo Treatment Effects on Lagged Earnings (LinkedIn vs. ACS)	245

1

General Introduction

1.1 Skills and Human Capital in the Labor Market

A large body of theoretical as well as empirical economics literature is devoted to identifying the causes of earnings differences between workers and countries. Pioneering work has recognized the stock of skills and other characteristics determining productive capability—that is, a worker’s *human capital*—as a key determinant (Mincer 1958; Mincer 1974; Becker 1962; Becker 1964). Human capital theory formalizes the notion that individuals deliberately invest in skills to raise their productivity and, therefore, earnings in the labor market (e.g., Schultz 1961). These benefits must be weighed against the direct and opportunity costs of such investments, including foregone earnings (Ben-Porath 1967). Empirically, differences in human capital or skills account for a substantial share of earnings variation both across countries (Hanushek and Woessmann 2008; Hendricks and Schoellman 2018) and within them (Deming 2022). Beyond descriptive evidence, a large body of empirical work emphasizes the causal nature of the relationship between human capital and earnings (e.g., Card 1999). This dissertation adds to this literature with four empirical contributions, focusing on the measurement, formation, and returns to skills acquired both in school and on the job.

Quantifying the importance of human capital for aggregate production and individual labor-market performance requires accurate measurement. Moving beyond early efforts that relied on literacy or enrollment rates, years of schooling and work experience have long served as workhorse proxies for the stock of human

capital (Mincer 1974; Heckman et al. 2006b). However, while these measures capture the quantity of education—inputs into the production of skills—they fail to directly measure the skills themselves. Several papers demonstrate that incorporating measures of cognitive skills—which reflect the quality of education—strengthens the observed relationship between human capital and economic growth (e.g., Hanushek and Woessmann 2012).

Accurate measurement also requires recognizing that the skill set relevant for productive tasks in the labor market is inherently multidimensional (Woessmann 2025; Deming and Silliman 2024). At the individual level, both cognitive skills, such as math and literacy, and non-cognitive skills, such as personality traits and social skills, are strongly associated with earnings (Hanushek et al. 2015; Edin et al. 2022). On the labor-demand side, analyses of job advertisements show that employers seek a broad range of skills tied to specific job tasks (e.g., Deming and Kahn 2018). Chapter 2 contributes to this literature by establishing self-reported skills from LinkedIn profiles as a measure of multidimensional human capital. We show that workers who report more skills tend to hold higher-paying jobs, and that these skills explain a much larger share of earnings variation than even detailed information on education and work experience.

Returns to skills are determined by the equilibrium of their supply and demand (Katz and Murphy 1992; Goldin and Katz 2007; Acemoglu and Autor 2011). Skill supply is shaped by individuals' decisions about how much and what type of education or training to acquire, while skill demand depends on the technology of production used in the workplace. This framework helps economists understand how skills relate to the distribution of earnings and how technological change affects the relative price of skills (e.g., Autor and Dorn 2013). Models formalizing this idea readily explain long-run trends in earnings inequality, such as the rising college wage premium observed over the last four decades (Autor et al. 2020).

Similarly, lower levels of schooling and work experience among women explained roughly a quarter of the gender earnings gap in the 1980s (Olivetti et al. 2024; Goldin 2024). Today, however, women outperform men in educational attainment. In addition, the gap in work experience has narrowed considerably, raising the question of what drives the persistent earnings gap between genders. Although differential sorting into industries and occupations now accounts for much of the remaining gap, this explanation is difficult to interpret. Chapter 2 contributes to this literature by showing that differences in skill sets between genders—beyond what is captured

by years of schooling or work experience—explain a substantial share of the gap, far exceeding the explanatory power of education or experience alone.

Human capital theory also addresses both the timing of educational investments and the types of skills acquired (Ben-Porath 1967; Sanders and Taber 2012; Becker 1962). Investing early allows for a longer period to reap the benefits. In addition, skills are often self-reinforcing over time, and some—such as cognitive skills—are most effectively developed early in life (Cunha and Heckman 2007). Although the time devoted to learning rather than earning typically declines after formal schooling, skills acquired on the job contribute just as much to the overall stock of human capital as those obtained through formal education (Jedwab et al. 2023; Black et al. 2023). Regarding the nature of skills acquired, formal education tends to provide more general, broadly applicable skills. By contrast, on-the-job learning allows workers to gain more specific skills tied to the tasks performed in their jobs (Becker 1964; Acemoglu and Pischke 1999b).

The success of educational investments in generating human capital depends on the production process of the underlying skills. This concept is formalized in economics through the skill production function, which relates educational outputs to the inputs used in producing skills (e.g., Hanushek 1986; Cunha and Heckman 2007). Inputs related to the organization and governance of school systems have received particular attention (e.g., Woessmann 2016). Differences in those inputs—such as whether students are tracked into different-ability schools or whether external exit exams are used—are able to explain cross-country differences in student achievement. Chapter 3 of this dissertation contributes to this literature by examining a specific rule school systems may adopt to evaluate students. In particular, we study whether grading students' comportment at school serves as a meaningful input in the production of non-cognitive and reading skills, and whether these grades facilitate a smoother school-to-work transition. Such effects could arise either through actual skill development or through signaling these skills to employers (Spence 1973). However, we find no evidence that comportment grades affect students' skill formation or their school-to-work transition.

Investing in one's own human capital is a decision made under uncertainty about future outcomes of that investment. Traditionally, economists have assumed specific objectives—such as income maximization—and used data on realized choices to infer the parameters of individuals' utility functions, such as the valuation of non-pecuniary aspects of work (e.g., Maestas et al. 2023). However, this approach

is problematic because observed choices can be rationalized by many different combinations of preferences and beliefs (Manski 2004). Since decisions under uncertainty are based on ex-ante beliefs, a better understanding of this decision-making process requires elicitation of individuals' subjective beliefs about the expected costs and benefits of their choices. A growing literature documents that such beliefs are key determinants of human capital investments (e.g., Arcidiacono et al. 2020; Wiswall and Zafar 2021). Chapter 4 contributes to this literature by eliciting workers' subjective beliefs about the pecuniary and non-pecuniary returns to firm-provided training, the most prevalent form of on-the-job learning. We show that workers perceive substantial returns in both dimensions, and that these beliefs are systematically related to both realized and intended training participation. To elicit returns beliefs, we rely on hypothetical investment scenarios (e.g., Dominitz and Manski 1996; Attanasio and Kaufmann 2014).

Causally identifying the career returns to skills is challenging because individuals choose how many and which skills to acquire. This selection is at least partly driven by unobserved factors—such as innate ability—that are also conducive to career success. To address this, economists have exploited natural experiments that generate random variation in schooling duration or admission to specific fields of study, allowing them to obtain credible estimates of returns to these types of education (e.g., Card 1999; Kirkeboen et al. 2016). Skill acquisition on the job, however, is less structured and considerably more heterogeneous, making it far more difficult to find exogenous sources of variation. Moreover, few data sources combine information on career trajectories with repeated observations of workers' skills. Chapter 5 addresses this gap by extending the data used in Chapter 2 to track longitudinal variation in self-reported skills on online professional profiles. By comparing workers with nearly identical career and earnings trajectories prior to reporting additional skills, I find that acquiring and reporting new skills is associated with higher subsequent earnings and increased upward job mobility into higher-paying positions.

1.2 Data

This section introduces the data sources employed in this dissertation. While Chapters 2 and 5 exclusively use career information from publicly accessible LinkedIn profiles of college graduates in the US, Chapter 4 relies on worker-level

data from a self-administered survey. Finally, Chapter 3 combines several sources of household- and individual-level survey data, student assessment studies, and a self-administered survey among teachers in Germany.

LinkedIn profiles are structured like resumes, listing employment histories, educational credentials including fields of study, and self-reported skills. These data are provided by Revelio Labs, a workforce intelligence company that compiles scraped profiles in a database. Revelio Labs enriches the data with further information such as occupation codes, earnings predicted from job characteristics, and gender predicted from first names, relying on proprietary algorithms. Chapter 2 extracts a cross-sectional snapshot of profiles from 8.85 million employed college graduates in the US in 2019. These are used to investigate whether self-reported skills are valid measures of human capital and can help rationalize commonly found patterns in the distribution of earnings. Chapter 5 builds a panel of these workers spanning the years between 2015 and 2023. The panel with 5.1 million college graduates is used to investigate to what extent longitudinal variation in workers' skills is related to downstream career outcomes. To enhance the robustness of the findings using predicted earnings, both Chapters 2 and 5 use occupation-level average earnings from the American Community Survey (ACS) as an alternative measure of earnings.

Eliciting workers' beliefs about returns to firm training is infeasible with administrative labor market datasets. Therefore, Chapter 4 relies on data from a self-administered online survey among 3,701 workers in Germany aged 25 to 55. The sample is representative in terms of age, gender, and geography and elicits detailed information about past as well as future intended firm training participation.

Chapter 3 combines repeated cross-section data on different student outcomes to investigate whether they are causally related to spatio-temporal variation in the receipt of compartment grades in Germany. More specifically, the Programme for International Student Assessment (PISA-E) data is used to measure students' literacy from standardized assessment tests. In addition, measures of non-cognitive skills are taken from the German Socio-Economic Panel (SOEP), and report card information comes from the Starting Cohort 3 from the National Educational Panel Study (NEPS SC3). Data on students' school-to-work transition is taken from the German Mikrozensus, a household survey covering 1% of the population with mandatory participation. Finally, we ran a survey among teachers in Germany to get a better sense of how compartment grading is implemented in practice.

1.3 Identification, Estimation Methods, and Overview

Empirical labor economics in the last few decades has emphasized two complementary tasks: identifying as well as estimating causal effects (e.g., Angrist and Pischke 2009), and improving the measurement of quantities that are central to theoretical economic models, often using novel sources of data. Each chapter of this thesis advances one or both of these tasks, and the next sections detail how.

On causal identification, a prime example that is central to this dissertation is the effect of human capital—specifically, skills—on career outcomes such as earnings and job mobility. Simple correlations between skills and earnings are insufficient to establish causality because skills are not randomly assigned. For instance, causality may run the other way: if higher-paying jobs are more conducive to skill accumulation, they will generate a positive cross-sectional association between skills and earnings, even if skills do not themselves raise earnings in this case.

To make progress, economists seek settings where the acquisition of skills is random to a certain degree. Random assignment permits a causal interpretation of statistical estimands like a (partial) correlation. Depending on the nature of this randomness, different methodological approaches can be employed to estimate and draw inference about a causal target parameter like the average treatment effect. In settings where random variation is scarce, researchers aiming to establish causality can instead work to rule out prominent alternative explanations that could generate similar patterns in the data.

Beyond causal inference, economists care deeply about measuring quantities central to theoretical models. If key objects are measured with error, conclusions from empirical analyses are misleading—if not meaningless.

Chapter 2 argues that better measurement of human capital—using self-reported skills on workers' online professional profiles—enhances our understanding of how skills relate to well-known labor market patterns, such as the gender earnings gap. We group more than 10,000 distinct raw skills into 48 skill clusters, which we assign to one of three skill classes: general, occupation-specific, and managerial. To establish the validity of our measure, we initially show that numbers of skills reported as well as their composition are systematically related to human capital investments such as education and work experience. In particular, higher actual work experience is associated with larger numbers of skills, and more advanced degrees provide more specific skills. Regarding the relationship between skills

and earnings, we show that workers with larger numbers of skills as well as more specific and managerial skills hold more highly-paid jobs.

In addition, we show that our skill measures explain more earnings variation than detailed measures of educational trajectories and work experience. To this end, we employ linear regressions and random forests (e.g., Breiman 2001) to optimize the predictive performance of our skill measures.

Finally, we document a sizable gender gap in both the number and composition of reported skills. We attribute slower skill growth among women to lower numbers of working hours associated with motherhood. Skill differences across gender can explain a substantial share of the gender gap in predicted earnings. In particular, accounting for skills yields larger reductions in the gap than adjusting for differences in education and work experience. To estimate this covariate-adjusted mean difference between male and female earnings, we rely on linear regressions as well as the causal forest algorithm introduced in (Athey et al. 2019). The chapter is joint work together with David Dorn, Moritz Seebacher, Lisa Simon, and Ludger Woessmann.

Chapter 3 exploits variation in the use of compartment grades across German federal states and calendar years to identify their effects on students' cognitive and non-cognitive skills, as well as on the school-to-work transition. The variation gives rise to a staggered difference-in-differences design. In each period, we compare students in states that have already implemented compartment grades to students in states that have not yet adopted them. Given the absence of pre-treatment differences in outcomes between treated and control groups, this comparison lends a causal interpretation to the effect of compartment grading on student outcomes. Our estimation strategy follows recent advances that caution against naive two-way fixed-effects estimation in staggered settings (e.g., Goodman-Bacon 2021; Callaway and Sant'Anna 2021). We find no effects of compartment grading on student outcomes and discuss the magnitudes that can be ruled out by our analyses. To probe this null finding, we use report-card data to show that compartment grades add little information once subject grades are known. In addition, we present survey evidence from teachers suggesting that, in the absence of compartment grading, schools substitute toward alternative pedagogical disciplinary measures. This chapter is joint work together with Larissa Zierow and Lukas Mergele.

Chapter 4 investigates how workers' beliefs about the returns to firm training shape their participation in it. We study workers' perceived causal effect of

participating in firm training on downstream career outcomes such as earnings, promotions, and job security. Because random variation in training participation is scarce, we overcome selection bias by eliciting perceived career outcomes using hypothetical scenarios (e.g., Dominitz and Manski 1996) in a within-subject design. We ask respondents to imagine two otherwise identical states of the world—with and without training. Therefore, any difference between scenarios can be attributed to firm training. We find that workers perceive substantial pecuniary and non-pecuniary returns to firm training. Lower-skilled workers expect lower non-pecuniary returns, partially explaining their lower uptake. In addition, we randomly assign an information treatment to estimate the causal effect of information provision on intended training participation (Haaland et al. 2023). Information provision increases training intentions primarily among lower-skilled workers, suggesting that addressing worker beliefs can help shrink the uptake gap between lower- and higher-skilled workers. This chapter is joint work with Hanna Brosch and Philipp Lergetporer.

Chapter 5 complements the evidence presented in Chapter 2 by exploiting longitudinal variation in workers' skill sets on LinkedIn. Building on a selection-on-observables assumption (Rosenbaum and Rubin 1983), I show that reporting additional skills on LinkedIn profiles within a more than two year long time window is related to career progression. In particular, in the three years after reporting additional skills (i.e., the treatment), I find gross yearly earnings increases of between 0.75 and 0.94% relative to the control group. Crucially, this finding also holds when using occupation-level average earnings from the ACS. In addition, workers reporting additional skills show higher job-mobility and are more likely to switch to a higher-paying position. To bolster the causal interpretation of these associations, I only compare workers with highly similar resumes and earnings trajectories before reporting additional skills in a linear regression. To ensure overlap in covariate distributions, I trim the sample based on estimated propensity scores. Due to the high dimensionality of the data, I use a flexible generalized random forest algorithm (Athey et al. 2019) to estimate propensity scores. Finally, I use placebo tests to bolster the causal interpretation of my findings (e.g., Imbens and Xu 2024). In particular, I show that there is no effect of reporting additional skills on lagged earnings – confirming an implication of the selection-on-observables assumption I employ to identify the effect of skills on career outcomes.

2

Multidimensional Skills on LinkedIn Profiles: Measuring Human Capital and the Gender Skill Gap¹

2.1 Introduction

Theories of human capital posit that individuals can invest in the acquisition of valuable labor-market skills. Workers enhance their earnings potential both by spending time in formal education and through on-the-job learning (Becker 1964, Mincer 1974). While measures of investments into human capital production such as years of schooling or years of work experience are available in many large-scale, worker-level data sets, the skills that result from such investments are usually unobserved. Education and experience variables are thus often used to proxy for the skills that result from such investments (Card 1999; Heckman et al. 2006b). However, task models of the labor market argue that workers with the same education level may possess a potentially large vector of distinct skills that qualifies them to perform different job tasks in the labor market (e.g., Autor et al. 2003). Analyses of job advertisements confirm that employers search for a wide range of worker skills that may not be easily captured by traditional education and experience metrics, such as problem-solving ability or proficiency with specialized software (e.g., Deming and Kahn 2018). There is, however, little direct evidence on which groups of workers

¹This chapter is based on co-authored work with David Dorn, Moritz Seebacher, Lisa Simon, and Ludger Woessmann. See Dorn et al. (2025) for the full reference.

possess the skills that employers value, and how such skills are related to workers' human capital investments and subsequent attainment of highly paid jobs.

In this paper, we provide large-scale evidence on the self-reported skill sets of millions of U.S. college graduates, using data from individual profiles on LinkedIn, the leading social media platform for professional networking. We conduct three sets of analyses that provide new evidence on skill differentials across distinct groups of college graduates and assess the usefulness of LinkedIn skills as a measure for human capital. First, we investigate whether skills systematically relate to education and experience as predicted by classical models of human capital acquisition. Second, we analyze whether the size and composition of workers' skill sets helps to explain which workers hold highly paid jobs. Third, we document skill differences between women and men, and relate these gender skill gaps to motherhood and to gender differentials in earnings.

LinkedIn profiles are structured like a worker's resume and provide information on self-declared skills as well as educational and occupational trajectories. Our analysis sample draws on a database of scraped public profiles curated by the workforce intelligence company Revelio Labs and contains a cross-section of LinkedIn profiles with complete skill, education, and experience data for 8.85 million college graduates who held jobs in the United States in 2019, before the onset of the Covid-19 pandemic.

To analyze the self-reported skills, we employ a word association algorithm to aggregate the reported raw skill strings into 48 skill clusters. We distinguish three broad groups of skill clusters. A first group comprises general skills that are used in a wide range of occupations, such as 'communication, problem solving' or 'customer satisfaction, customer retention'. A second group contains specific skills that are concentrated in a narrow set of occupations, such as 'legal research, legal writing' or 'SQL, software development'. And a third group consists of managerial skills that are frequently observed in top executive jobs, such as 'project management, budgets' or 'coaching, leadership development'. In our analysis, we discuss and address various measurement challenges emerging from the self-reporting of skills.

Our first set of results analyzes the relationship between skills and measures of experience and education. We document that the average age-skill profile has a concave shape which looks remarkably similar to the well-known average age-earnings profiles. Older workers report a larger number of skills than their younger peers, which is consistent with an acquisition of skills through on-the-job

experience. That interpretation is supported by the observation that the number of skills increases more strongly with actual rather than potential years of experience. It is further corroborated by the finding that older workers indicate larger fractions of occupation-specific and managerial skills, suggesting that these skills are often acquired through on-the-job experience, while general skills that are not focused on specific occupations are primarily acquired during initial education. The composition of reported skills also varies with different types of college education. Most notably, workers with degrees in arts, humanities, or social sciences have considerably higher shares of general skills than graduates of science, medicine, or law programs (who have relatively more specific skills) and graduates of business programs (who have relatively more managerial skills).

We next study the relationship between skills and the attainment of highly paid jobs. We draw on salary data that Revelio Labs imputes primarily based on job characteristics such as job title and firm name. We show that workers with a larger number of skills, and those who report higher fractions of managerial and occupation-specific skills, tend to have higher-paid jobs. The skill variables from LinkedIn account for a larger proportion of job-based earnings variation than an augmented Mincer model with a highly detailed vector of education and experience variables, and they mediate the impact of these traditional human capital variables to a substantial extent. We obtain qualitatively similar results with a simple alternative earnings metric that relies on data from the Census Bureau's American Community Survey and assigns to each worker the average earnings in her occupation.

Through the lens of these relationships between education, experience, skills, and earnings, we analyze gender gaps between women and men. Young women in their twenties report only a slightly lower average number of skills as their male peers, and this difference is fully accounted for by gender-specific occupational choice. But women's age-skill profiles are considerably flatter than men's during their thirties and early forties, giving rise to a substantial gender skill gap. We provide evidence that this emerging skill gap is unlikely to be primarily driven by gender differences in overreporting, profile updating, or occupational choice. Instead, the evidence is consistent with an explanation where reduced female labor supply due to motherhood gives rise to slower skill accumulation on the job. Gender differences in self-reported skills can account for 60 percent of the gender gap in job-based earnings, a proportion that considerably exceeds the explanatory power of highly detailed education and experience variables.

Taken together, our results suggest that worker skills on LinkedIn profiles, despite being self-reported and potentially noisy, provide an informative new measurement of individual human capital that is richer and more detailed than most available measurements.

The remainder of this paper is organized as follows. Section 2.2 relates our analyses to several strands of the existing literature. Section 2.3 describes the LinkedIn data and discusses potential challenges for the measurement of skills. Section 2.4 defines variables and our analysis sample of college graduates who work in the US labor market. Section 2.5 tests whether skills relate to experience and education as predicted by basic theories of human capital acquisition, while section 2.6 assesses whether more skilled workers are employed in higher-paid jobs. Section 2.7 documents a gender skill gap and investigates its relation to motherhood and to the gender earnings gap. Section 2.8 concludes.

2.2 Related Literature

Our analysis contributes to the large literature on skills in the labor market. Seminal human capital models characterize the accumulation of skills in formal education and on the job (Becker 1964; Mincer 1974; Heckman et al. 2006b). In a framework of supply and demand for skills, the distribution of earnings by skill level emerges from a race between an evolving supply of worker skills and advances in production technology that shape the demand for skills (Tinbergen 1974; Katz and Murphy 1992; Goldin and Katz 2007; Autor 2014; Autor et al. 2020). Task models posit the importance of multidimensional skills to perform a broad range of job tasks required in the labor market (e.g., Autor et al. 2003; Acemoglu and Autor 2011; Autor and Dorn 2013). However, in all these literatures skills are usually not directly observed but instead proxied by measures of education and work experience.² Data from online postings of job vacancies, however, show that employers are interested in a wide range of skills beyond educational credentials (e.g., Deming and Kahn 2018; Hershbein and Kahn 2018; Deming and Noray 2020; Braxton and Taska 2023). But the job advertisement data do not directly indicate which workers hold these

²In human capital models with a unidimensional skill and competitive labor markets (e.g., Mincer 1974), a worker's skill can be directly inferred from her earnings, and no separate measurement of skill is needed. In a setting with multidimensional skills (and possibly labor markets that are not perfectly competitive), it is no longer possible to simply back out a worker's skill set from her earnings.

multidimensional skills. Our analysis complements this literature by measuring multidimensional skills in worker-level data.³ In addition, we can observe a much richer set of education variables than most classical studies of human capital.

Prior analyses of individual skills usually observe subsets of skills and study their relationship with earnings (see Deming and Silliman 2024; Woessmann 2025 for recent reviews of the literature). Some investigate skills in various cognitive domains such as literacy, numeracy, and science (e.g., Murnane et al. 1995; Bowles et al. 2001; Hanushek and Woessmann 2008; Castex and Kogan Dechter 2014; Hanushek et al. 2015; Hermo et al. 2022). Others study social skills such as leadership and teamwork (e.g., Kuhn and Weinberger 2005; Weinberger 2014; Deming 2017; Deming 2023a) and other noncognitive aspects of skills (e.g., Heckman et al. 2006a; Lindqvist and Vestman 2011; Edin et al. 2022), partly in various combinations (e.g., Piopiunik et al. 2020; Lise and Postel-Vinay 2020; Guvenen et al. 2020). However, such studies tend to fall short of covering the wide range of skills seen in job advertisements or observed in our data. Prior use of LinkedIn data for labor-market analysis is limited (e.g., Wheeler et al. 2022; Conzelmann et al. 2022; Evsyukova et al. 2025) and does not make use of the skill information contained in the professional profiles.

Our categorization of skills into general versus specific skills follows a long tradition in labor economics (Streeck 2011). This distinction, a central aspect of the original Becker (1964) model, is particularly relevant when analyzing job tasks (Sanders and Taber 2012) and life cycle dynamics (Hanushek et al. 2017). General skills are typically defined as being portable, in the sense that they can be productively used in many different jobs, whereas specific skills are valuable only in a limited set of jobs. Becker (1964) adopted a narrow definition of specificity, where specific skills are useful only at a single firm. We instead focus on the occupational specificity of skills, consistent with the empirical observation that experience acquired at one firm often remains beneficial when workers move to a similar occupation with comparable tasks at another firm (Gathmann and Schönberg 2010; Eggenberger et al. 2018). A standard assumption in human capital models is that formal education provides primarily general skills, whereas specific skills are acquired relatively more through on-the-job experience (e.g., Acemoglu and Pischke 1999b).

³A measurement advantage in our data relative to job ads is that LinkedIn profiles comprise a dedicated field where users indicate skills. Job ads are instead often unstructured so that researcher discretion is necessary to define which elements of a job ad refer to skill requirements.

We additionally subdivide general skills into managerial and non-managerial skills. Managerial skills, which we define as the skills that are most concentrated among top executives, are general in the sense of having a high portability across occupations. However, the underrepresentation of recent college graduates in management jobs suggests that managerial skills are acquired more through work experience than through education, which makes them more comparable to specific skills in this regard. Our conceptualization of managerial skills is related to leadership skills (Kuhn and Weinberger 2005), people management skills (Hoffman and Tadelis 2021), and skills to coordinate, monitor, and motivate workers (Weidmann et al. 2024), but it is broader as it also includes skills related to tasks such as planning, budgeting, or business analysis that are not necessarily focused on people.

Human capital differences have long been considered as a potential explanation for differences in labor-market outcomes across different population segments, perhaps most notably the earnings gap between women and men (Goldin 2024). During the 1980s, about one quarter of the gender gap in full-time wages in the U.S. could be explained by women's lower average years of education and experience (Olivetti et al. 2024). Over the following three decades, the male advantage in work experience shrank and women overtook men in terms of educational attainment. By 2018, basic education and experience variables can no longer rationalize a gender wage gap in favor of men and instead predict a modest gap in favor of women.⁴ In the words of Goldin (2014, p. 1094), "as women have increased their productivity enhancing characteristics and as they 'look' more like men, the human capital part of the wage difference has been squeezed out."

While education and experience have lost explanatory power, the contribution of occupation and industry of employment to the gender wage gap has increased from 19 percent in 1980 to 54 percent in 2018 (Olivetti et al. 2024). Blau and Kahn (2017, p. 797) note that these occupation-industry effects "may represent human capital, other labor-market skills, and commitment, on the one hand, or employer discrimination, on the other hand", which raises the possibility that the contribution of skill differentials to the gender wage gap may be considerably larger than implied

⁴Using survey data from the Panel Study of Income Dynamics, Olivetti et al. 2024 estimate that education and experience explained 23 percent of the gender wage gap in 1980, 25 percent in 1989, 17 percent in 1998, 8 percent in 2010, and minus 8 percent in 2018.

by basic education and experience variables alone.⁵ We contribute to this literature by documenting a gender gap in self-reported skills, and by showing that the large and difficult-to-interpret occupation- or job-specific component of the gender earnings gap can to a substantial extent be rationalized by these skill differences.⁶

We also contribute to the literature examining the role of motherhood in the development of gender gaps over the life cycle (Bertrand et al. 2010; Kleven et al. 2019). Our results show that motherhood is associated with a reduction in women's work hours and a slower accumulation of skills compared to men of the same age.

Prior research on gender differences in skills has focused primarily on numeracy and literacy. While there are gender gaps especially in numeracy, they account for only a small portion of the gender earnings gap (Rebollo-Sanz and De la Rica 2022). The trajectory of numeracy and literacy skills with age is much more positive for individuals who use them regularly (Hanushek et al. 2025). However, the evidence is mixed on whether unemployment leads to substantial skill decline (Edin and Gustavsson 2008; Dinerstein et al. 2022; Cohen et al. 2025; Arellano-Bover 2022) and whether parenthood results in a differential erosion of skills for mothers relative to fathers (Jessen et al. 2025). Our research contributes to this literature by examining a much broader set of skills. It documents gender differences in the composition and accumulation of new skills, and suggests that these differences play a more important role in explaining the gender earnings gap than previously recognized.

2.3 LinkedIn Data and the Measurement of Skills

The primary source of our data consists of publicly accessible individual worker profiles obtained from the professional networking platform LinkedIn, which we introduce in section 2.3.1. We observe a cross-section of these profiles in 2019 and discuss in section 2.3.2 the limitations imposed by the cross-sectional approach. The LinkedIn profiles offer detailed self-reported information on workers' skills.

⁵Differences in occupational composition between women and men are partly driven by gender disparities in college major choices (Speer 2020; Bertrand 2018) However, Sloane et al. 2021 show that accounting for college major reduces the contribution of occupational sorting to the gender wage gap only modestly.

⁶A few studies have succeeded in nearly fully accounting for gender wage gaps within narrow groups of highly educated workers. Azmat and Ferrer (2017) show that among young lawyers, the earnings gap is largely explained by performance differences, while Bertrand et al. 2010 find that the gap among MBA graduates primarily reflects gender differences in prior training, career interruptions, and weekly work hours. However, it remains difficult to determine to what extent these factors reflect underlying gender differences in skills.

We discuss in section 2.3.3 how users report skills on their profiles and in how far self-reporting generates measurement challenges for our analysis.

2.3.1 LinkedIn Profiles

LinkedIn is the world's largest professional network and operates a social media platform that focuses on career development. Its more than one billion registered users worldwide (LinkedIn 2024) create online profiles that list CV information such as current and prior jobs, educational degrees, and skills. LinkedIn is widely used by workers and employers for job search and recruiting (Evsyukova et al. 2025), and many universities and career consultants recommend creating and regularly updating profiles on the platform (e.g., Kratz 2021; Arizona State University Alumni 2021). In the United States, a majority of college graduates reports using LinkedIn (Auxier and Anderson 2021), and most corporate recruiters look for job candidates on the platform (Ryan 2020).

Most of the CV information on LinkedIn is part of a public profile that is visible to any user of the platform, while other information such as posts and shares is usually only visible to users that are logged in on LinkedIn or who have established a personal link with the profile holder on the platform. The LinkedIn records analyzed in our study are provided by Revelio Labs, a workforce intelligence company that builds an encompassing human-resource database from various sources. Revelio Labs scrapes the semi-structured public profiles on LinkedIn and then parses and curates the information into a structured database. They clean, standardize, and enrich the data using proprietary algorithms. Our raw data consist of a cross-section of LinkedIn records from the United States that were scraped between January and September 2019, before the massive labor-market disruptions caused by the Covid-19 pandemic.

2.3.2 Cross-Sectional Analysis

The cross-sectional structure of the data does not allow us to distinguish between age and cohort effects when analyzing skill differences across younger versus older workers. Our descriptive analysis is thus similar in spirit to Mincer's (1974) classical analysis of the relationship between education, potential experience, and earnings. Mincer's pioneering work has subsequently been complemented by longitudinal studies which establish that the originally observed cross-sectional

patterns largely capture age effects and not just cohort effects (Thornton et al. 1997; Sandgren 2007) and by analyses with causal designs that establish a causal link between human capital investments and earnings (see Card 1999; Heckman et al. 2006b for reviews of the literature).

Consistent with human capital models that predict an accumulation of skills as workers age, and following a large empirical literature starting with Mincer (1974), we interpret the combined age-cohort effects in our data as age effects. This choice simplifies the language we can use to describe results but comes with the caveat that cohort effects can never be fully ruled out.⁷

2.3.3 Self-Reported Skills

Workers self-report skills on their LinkedIn profiles. The addition of skills to a user's profile is semi-structured. When the user starts typing a skill, LinkedIn provides suggestions of frequently used skill terms to facilitate comparability across user profiles. However, users are free to enter new skill terms that have not been previously used on the platform. We observe 1.8 million unique text strings for skills in our analysis sample, and the average profile that reports skills indicates 20 of them. Before describing our definition of skill variables and other relevant variables in section 2.4, we discuss potential measurement challenges that result from the self-reporting of skills and their implications for our empirical analysis.

Relevance of LinkedIn Skills. LinkedIn profiles are widely used for recruiting and job search, and skill information on users' profiles plays an important role in this process. In a survey of recruiting professionals who use LinkedIn, 94 percent of recruiters agree that it necessary to understand which skills workers possess to make informed talent decisions (Degraux 2023). According to LinkedIn, 48 percent of company recruiters on the platform explicitly use workers' skill data when they seek to fill vacancies (Anderson 2024), and searches by skill are considerably more likely than searches by years of experience (Degraux 2023). Whether or not a worker lists a particular skill thus affects the likelihood of being approached by a recruiter

⁷Cross-sectional analyses that relate earnings to potential or actual work experience remain widely used to measure human capital accumulation (e.g., Jedwab et al. 2023). An advantage of a cross-sectional analysis over a panel analysis is that the former does not confound age effects with year effects. Since 2023, LinkedIn phased in the feature that users can report skills separately for each position they held (Mason 2023). To the extent that this feature changed overall patterns of skill reporting, a panel analysis that follows individuals over time would tend to confound age effects with year-specific reporting effects.

(Smith 2023). LinkedIn skills should therefore be of interest to labor economists because they have direct relevance for labor-market outcomes. To relate these skills to theories of human capital, we however must consider to what extent workers' self-reports of skills correspond to a theoretically grounded notion of skills.

Definition of Skills. A leading dictionary defines skills as "a learned power of doing something competently" (Merriam-Webster 2025), which is consistent with the premise of human capital theory that skills are produced through a learning process in school or on-the-job. In our empirical analysis, skills are however potentially more broadly defined as any concept that LinkedIn users report in the skill section of their profiles. While it is not guaranteed that users' understanding of skills always follows the dictionary definition, we show in section 2.4.1 that the most frequently reported skill strings indeed correspond to aptitudes that have plausibly been acquired in a learning process. We also note that there is substantial overlap between the keywords that researchers have used to extract skill requirements from job ads (e.g., Deming and Noray 2020) and the skills most frequently reported on LinkedIn profiles.

Incomplete and Outdated LinkedIn Profiles. LinkedIn data will underreport skills (and other relevant profile information) if users never created a complete profile or ceased to update their profile. We address these challenges as follows. First, our sample selection process described in section 2.4.4 removes apparently incomplete profiles, including those that report no skills or no educational information. Second, we present robustness tests in Appendix B for the subsample of workers who moved to a new employer within the last two years. The indication of a new employer ascertains that these workers modified their profile relatively recently, and the preceding job search arguably created an incentive to update and complete the entire profile information. While the subsample of recent movers should include user profiles with higher information quality, its focus on workers with recent job mobility makes it less representative of the overall population of college graduates in the US labor market, and we thus use it for robustness tests rather than for our main analysis. Appendix B shows that the qualitative results of our main analyses are confirmed in the sample of recent firm switchers, indicating that these results are unlikely to be driven by underreporting in outdated profiles.

Underreporting of Basic Skills. Workers typically use their LinkedIn profiles to highlight information that differentiates them from their peers while omitting widely shared qualifications. For example, college graduates rarely provide a full record

of their earlier education, such as elementary or high school, nor do they mention basic skills like reading. Similarly, users who list advanced competencies, such as proficiency in specialized data analysis software, may leave out more common abilities, such as using basic spreadsheet programs. If highly skilled individuals are more likely to underreport basic skills, such reporting bias may dampen an association between human capital investments and the number of reported skills, and between the number of skills and workers' earnings. This bias would work in the opposite direction of the results we find in our subsequent analysis.

Idiosyncratic Reporting of Skill Detail and Proficiency. LinkedIn users may apply different idiosyncratic standards to evaluate their own skill sets. First, users can decide on the level of detail of the skills they want to indicate. Whereas one user may report to be skilled in the 'MS Office' software suite, another user may indicate mastery of the individual programs 'MS Word' and 'MS Excel' as two separate skills. Second, users may differ in their assessment of how high one's level of proficiency should be to claim being skilled in domains such as 'project engineering', 'public speaking', or 'software development'.⁸ Some users may even knowingly state skills that they do not possess to attract the attention of recruiters, even though career advisory services warn that such a strategy could backfire later in the recruiting process (UCLA Career Center 2024).

The observation that many recruiters use the LinkedIn skill data to search for suitable job candidates suggests that professional users of the platform consider the skill data to be generally informative about workers' actual qualifications. LinkedIn also seeks to enhance the quality of the reported skill information with a feature that allows people to endorse the skills reported by other users (typically co-workers) to whom they are connected on the platform. While this endorsement information is not part of the public profiles that can be scraped and is thus not contained in our data, the social control exerted by other users likely helps to improve the quality of the skill data that we have available.⁹

⁸Idiosyncratic reporting practices presumably introduce less noise in the measurement of other relevant variables on LinkedIn profiles such as education or work experience, since the question whether a worker has obtained an educational degree or held a job is less subject to individual interpretation. The reporting challenges that we discuss here for skills reported on LinkedIn profiles however also apply to the more established literature that studies the skill requirements reported in job advertisements, in which companies may apply different practices to report such skill demands.

⁹Users also have the option to report profiles with inaccurate information directly to LinkedIn, which will then review and potentially ban the profile.

For the purpose of our empirical analyses, the idiosyncratic reporting of skill detail and proficiency implies that the skill information contained in individual profiles provides a noisy measure of the skill sets that one would see if skills were assessed based on a common, objective standard. Noise in the measurement of individual-level skills will create attenuation bias when we regress job-based earnings on individuals' skills. For most parts of our subsequent analyses, we however aggregate the skill information of individuals to large groups of workers delineated by such variables as age, education level, or gender, and then focus on average skill differences across these groups. If the distribution of reporting noise is the same within each such group (e.g., if workers of different ages are equally likely to report a given set of skills using fewer or more skill strings), then the skill differences we measure across groups will be the same that would have been observed with a hypothetical objective measure of skills.¹⁰ The group-level comparisons will be biased only if the groups have a different reporting behavior on average. While we do not see a strong reason to expect differential group-specific reporting for many of the worker groups we study, we will address in section 2.7.1 whether observed skill gaps between women and men might be the result of gendered reporting behavior.

2.4 Definition of Variables and Sample

We next describe how we organize the detailed skills reported on LinkedIn profiles into skill clusters and how we classify these clusters into the three broad domains of general non-managerial, general managerial, and specific skills (section 2.4.1). The following sections define other relevant variables for our analysis that are either directly observed in the profiles (section 2.4.2) or inferred from other profile information (section 2.4.3). Section 2.4.4 outlines our sample selection process and evaluates the representativeness of the final dataset.

¹⁰Suppose a worker i of group j has s_{ij} skills according to an unobserved objective skill assessment, but she reports $\sigma_{ij} = s_{ij} + \varepsilon_{ij}$ skills on her LinkedIn profile, where $\varepsilon_{ij} \sim F(\mu_j, \theta_j)$ is a noise term drawn from a distribution function $F(\cdot)$ with group-specific mean μ_j and shape parameter θ_j . Due to the law of large numbers, the comparison of group means in reported skills between two groups $j=1$ and $j=2$ will tend towards $\bar{\sigma}_1 - \bar{\sigma}_2 \rightarrow \bar{s}_1 - \bar{s}_2 + (\mu_1 - \mu_2)$, and provides an unbiased measure of the objective skill differential $\bar{s}_1 - \bar{s}_2$ if $\mu_1 = \mu_2$.

2.4.1 Skill Clusters

Derivation of Skill Clusters. We observe 1.8 million unique text strings for skills in our analysis sample of LinkedIn data, which retains profiles that report at least one skill (see section 2.4.4 below for more details on sample selection). The average such profile lists 20 skills, while the numbers of skills at the 10th and 90th percentiles of the distribution are 7 and 37, respectively.

To reduce the dimensionality of the skill data, we group skills that frequently appear together in the same user profiles into skill clusters. Revelio Labs cleans the skill terms and uses a Word2Vec word association algorithm to cluster the most common 10,500 skill strings (which account for more than 90 percent of all skill entries on U.S. LinkedIn profiles) into 50 clusters based on their co-occurrence patterns. After disregarding three clusters that report skill terms in languages other than English¹¹ and adding a residual cluster that aggregates all infrequent skill strings that are not among the 10,500 most common ones, we retain 48 skill clusters for our main analysis.

Table 2.1 lists the 48 skill clusters, which are named according to the two most frequent skill strings contained in the cluster. Appendix Table A1 provides further information on cluster composition by listing each cluster’s five most frequent skill strings. We observe that the clustering algorithm usually groups terms that refer to either adjacent or overlapping skill concepts. For example, the five most frequent skill strings in the cluster ‘accounting, financial reporting’ are ‘accounting’, ‘financial reporting’, ‘auditing’, ‘financial accounting’, and ‘accounts payable’.

Table 2.1: Skill clusters: Descriptive statistics

	Mean (1)	Share positive (2)	Specificity (3)	Executive (4)
Specific skills				
Accounting, financial reporting	0.409	0.094	0.604	−0.005
AutoCAD, SolidWorks	0.166	0.062	0.733	−0.031
Biotechnology, molecular biology	0.210	0.042	0.656	−0.012
Clinical research, medical devices	0.222	0.070	0.650	−0.013
Engineering, project engineering	0.289	0.086	0.618	−0.011

¹¹Three of the 50 skill clusters group a panoply of skills that were entered in Spanish, French, or Portuguese instead of English. Our analysis omits profiles with these skill clusters, since they cluster skills by the languages used to describe the skills rather than the content of the skills.

Table 2.1: Skill clusters: Descriptive statistics (continued)

	Mean (1)	Share positive (2)	Specificity (3)	Executive (4)
Healthcare, hospitals	0.499	0.099	0.629	-0.011
Insurance, banking	0.266	0.066	0.610	0.013
Java enterprise edition, Jira	0.104	0.033	0.780	-0.025
Java, Matlab	0.397	0.130	0.624	-0.067
Legal research/writing	0.337	0.049	0.732	0.030
Mobile devices/applications	0.047	0.028	0.611	0.002
Real estate, investment properties	0.191	0.039	0.644	0.046
Revit, SketchUp	0.168	0.051	0.640	0.001
SQL, software development	0.515	0.109	0.704	-0.053
Telecom., network security	0.169	0.043	0.646	-0.006
Windows server, disaster recovery	0.134	0.045	0.698	-0.004
General managerial skills				
Analysis, financial analysis	0.579	0.176	0.501	0.106
Business analysis/process improv.	0.351	0.128	0.515	0.068
Coaching, leadership development	0.477	0.147	0.455	0.071
Marketing, social media marketing	0.741	0.223	0.462	0.110
Program mgmt., security clearance	0.349	0.115	0.407	0.084
Project management, budgets	0.636	0.285	0.279	0.151
Sales, strategic planning	1.325	0.408	0.279	0.238
General non-managerial skills				
Access, software documentation	0.131	0.080	0.515	-0.038
Communication, problem solving	0.238	0.135	0.224	-0.036
Customer satisfaction/retention	0.113	0.073	0.418	-0.009
Data analysis, databases	0.363	0.162	0.398	-0.059
Editing, public relations	0.525	0.174	0.467	0.036
Energy, sustainability	0.148	0.045	0.549	0.025
English, Spanish	0.166	0.108	0.214	-0.014
Excel, customer relations	0.075	0.051	0.261	0.004
Food, hospitality	0.191	0.046	0.443	0.009
HTML, JavaScript	0.311	0.099	0.581	-0.038
Info. technology, lean six sigma	0.045	0.038	0.404	0.016
Inventory/operations management	0.270	0.110	0.476	0.050
Microsoft Office, customer service	2.861	0.682	0.127	0.020
Other	1.382	0.422	0.103	-0.025

Table 2.1: Skill clusters: Descriptive statistics (continued)

	Mean (1)	Share positive (2)	Specificity (3)	Executive (4)
Photoshop, Adobe CS	0.586	0.154	0.435	−0.052
Process improv., cross-func. team lead.	0.397	0.146	0.486	0.036
Public speaking, research	1.406	0.405	0.284	0.030
Recruiting, human resources	0.315	0.092	0.445	0.022
Retail, forecasting	0.322	0.113	0.471	0.018
Security, emergency management	0.156	0.048	0.488	0.008
Social media/networking	0.498	0.242	0.363	−0.014
Testing, quality assurance	0.158	0.091	0.507	−0.034
U.s, software dev. life cycle ^a	0.034	0.031	0.390	−0.003
Video production/editing	0.379	0.080	0.523	−0.009
Windows, troubleshooting	0.269	0.109	0.417	−0.040

^a The abbreviation “U.s” in “U.s, software development life cycle” likely stands for “user stories,” a feature of agile software development aimed to represent users’ requirements. This interpretation is consistent with other frequent strings in this cluster, such as “relationship building” and “waterfall methodologies,” which also reflect customer-oriented software development skills. *Notes:* Column 1 reports the average number of raw skills reported in the cluster. Column 2 shows the share of profiles with at least one raw skill reported in the cluster. Column 3 indicates the Gini coefficient of skill concentration across occupations. Column 4 reports the difference between the fractions of profiles in the “top executives” occupation group and in all other occupation groups that report at least one skill in the cluster. Sample: U.S.-based workers with a college degree whose LinkedIn profiles contain required information. $N = 8,850,314$.

The skill dimensions emerging from the clustering algorithm paint a rich picture of the multidimensional skills reported by workers. Many skill clusters refer to specific functions within firms such as ‘recruiting, human resources’, ‘project management, budgets’, or ‘customer satisfaction, customer retention’, or they group broadly applicable skills such as ‘communication, problem solving’ or ‘Microsoft Office, customer service’. Other skill clusters instead indicate expertise related to a specific field of knowledge such as ‘clinical research, medical devices’ and ‘legal research, legal writing’, or show familiarity with software and technology applications such as ‘Photoshop, Adobe Creative Suite’ or ‘mobile devices, mobile applications’.

While LinkedIn users are in principle free to list any concept as a skill on their profiles, the list of most frequent skill strings by cluster in Appendix Table A1 suggests that users typically understand skills as learned abilities that may plausibly result from education, training, and on-the-job experience. The skills

indicated by workers on LinkedIn also show considerable overlap with widely used skill terms in job ads.¹²

Classification of Skill Clusters by Occupational Specificity. We characterize the occupational specificity of each skill cluster based on the extent to which these skills are concentrated among workers of a few occupations rather than being widespread across occupations. For each LinkedIn profile, we measure whether it lists at least one skill of a given skill cluster and then compute each cluster's Gini coefficient across 336 six-digit occupations based on the Standard Occupational Classification (SOC). The resulting specificity measure for each skill cluster is reported in column 3 of Table 2.1, and we classify the top tercile of skill clusters with the highest Gini coefficients as *specific skills*. These specific skills include knowledge of specialized software for applications such as computer-aided product design ('AutoCAD, SolidWorks') or architectural planning ('Revit, SketchUp'), as well as skills that are specific to areas such as medicine/health ('clinical research, medical devices', 'healthcare, hospitals'), law ('legal research, legal writing'), finance ('insurance, banking', 'accounting, financial reporting'), and science/engineering ('biotechnology, molecular biology', 'engineering, project engineering').

We refer to the two terciles of skill clusters with lower Gini measures of occupational concentration as *general skills*. Many of these skills relate to basic business functions in firms, sometimes combined with the use of generic workplace software.

Classification of General Skill Clusters into Managerial and Non-Managerial Skills. We observe that general skills comprise relatively basic skills such as 'Microsoft Office, customer service' or 'editing, public relations' that may qualify workers for entry-level jobs in firms, but also more advanced skills such as 'project management, budget' or 'coaching, leadership development' that would seem suitable for managerial positions. To reflect this distinction, which is relevant for some of our empirical results below, we further subdivide general skills into general managerial and general non-managerial skills. To this end, we compute

¹²Data Appendix Table 2.1 of Deming and Noray (2020) lists 51 common skill strings in job ads which they group to the broader domains of social, cognitive, creativity, writing, management, finance, business systems, customer service, office software, technical support, data analysis, and specialized software skills. Nearly half of these skills (24 of 51), and at least one per domain, also appear among the frequently used skill terms on LinkedIn profiles shown in Appendix Table A1. We find little coverage in the LinkedIn data of the domain that Deming and Noray 2020 describe as character, which comprises keywords such as 'detail-oriented', 'energetic', or 'self-starter' that may less clearly relate to the concept of skills as learned abilities. We also do not find keywords related to machine learning and artificial intelligence among the top skills in Appendix Table A1, perhaps because these technologies were not yet as broadly used in 2019.

for each skill cluster whether the corresponding skills are more concentrated in the occupation group ‘top executives’ (code 11-1000 of the 2018 SOC occupational classification) than among all other occupations.¹³ The seven skill clusters that are most concentrated among top executives are classified as *general managerial skills*, with the remainder of the general skills being *general non-managerial skills*. It is noteworthy that the skill clusters with the highest concentration among top executives all qualify as general skills as per our definition above. While these skills are particularly frequent among top executives, they are observed across a broad range of other occupations. This pattern likely results because managerial occupations capture only a small subset of the workers who perform managerial functions across many different occupations.

Our approach of classifying all skill clusters into the three categories specific skills, general managerial skills, and general non-managerial skills is both comprehensive and rule-based. It differs from other analyses that classified the skills of workers or the skill requirements of jobs as routine/non-routine (e.g., Autor et al. 2003; Spitz-Oener 2006; Autor and Handel 2013) or social/cognitive/non-cognitive (Deming 2017) based on only a carefully selected subset of available skill or task variables. The advantage of our approach is that we can include all skills that workers declare in our analysis since occupational specificity is a well-defined concept that can be applied to every skill. The rule-based classification approach reduces researcher discretion but can occasionally generate a classification that does not fit researcher intuition perfectly. For instance, the skill cluster ‘sales, strategic planning’ whose five most frequent skill strings are ‘sales’, ‘strategic planning’, ‘team leadership’, ‘account management’, and ‘strategy’ is classified as a managerial skill, even though ‘sales’ alone might intuitively be seen as a general non-managerial skill.

Table 2.1 provides descriptive statistics for the frequency of skills in our sample, as well as their overall occupational specificity and their concentration among top executives. Column 1 of the table indicates that the average LinkedIn profile in our sample comprises more than one skill string from each of the three skill clusters ‘Microsoft Office, customer service’, ‘public speaking, research’, and ‘sales, strategic planning’, while the four least frequent clusters contribute less than 0.1 skills each to the average profile.

¹³Specifically, we compute the fraction of profiles in the ‘top executives’ occupation group that report at least one skill from a given skill cluster and subtract from this the fraction of profiles from all other occupation groups that include skills from this cluster. The resulting statistic is indicated in column 4 of Table 2.1.

There is a strong negative correlation (of -0.53) between skill frequency and skill specificity, which means that many specific skills are also relatively rare. While we classify a third of all skill clusters as specific skills, only one-fifth of the skills on the average LinkedIn profile are specific (4.1 out of 19.9 skills). Conversely, the seven general managerial skills are relatively frequent, accounting for slightly over one fifth of the skills on the average profile (4.6 out of 19.9 skills).

2.4.2 Additional Variables Observed in LinkedIn Profiles

Education. LinkedIn profiles usually report educational degrees along with a field of study and the name of the educational institution that granted the degree. Our data cleaning seeks to standardize this information to obtain three variables: highest degree, field of study for the highest college degree, and college quality as proxied by a university ranking.

We distinguish six levels of education degrees: high school, associate, bachelor, master, professional degree, and doctorate. We parse the educational information from LinkedIn into these categories using information from several online sources on typical abbreviations of degrees (StudentNews Group 2024; Best Universities 2024; YourDictionary 2022) and subsequently hand-map the most frequent remaining categories. The most common non-mapped entries are either missing values or ambiguous terms such as ‘certificate’ or ‘diploma’. If users report more than one college degree, we determine the highest degree based on the following descending ranking: doctorate, professional degree, master, bachelor, associate. If they report multiple degrees of the same type, we use the most recent one.

To classify the field of study of the highest degree, we start with the Classification of Instructional Programs (CIP) taxonomy from the National Center for Education Statistics (2024). As LinkedIn’s field of study entries are semi-structured, most users report a field of study that corresponds to a CIP instructional program. For the largest non-mapped field-of-study entries, we again use a manual mapping. The most common field-of-study entries that remain unmapped are missing values, GPAs, or imprecise terms such as ‘science’. We proxy for educational quality by merging information from the Times Higher Education (2019) U.S. College Ranking to each user profile based on the name of the college where the highest degree was obtained.

Experience. LinkedIn profiles usually provide graduation dates for their educational degrees and start and end dates for each job spell. We use this information to

calculate the actual work experience that every individual accumulated between graduation from the first undergraduate degree (associate or bachelor) and the scrape date in 2019.¹⁴ We also compute potential experience as the entire time interval between graduation and scrape date.

Occupation and Location. LinkedIn entries for jobs typically contain information on occupation and place of work.¹⁵ Revelio Labs aggregates the occupational information to 336 occupation groups according to the 2018 SOC classification and codes the place of work based on Metropolitan Statistical Areas (MSA) and U.S. states.

2.4.3 Variables Inferred from LinkedIn Profiles and External Sources

Gender. LinkedIn profiles do not usually indicate workers' gender. Revelio Labs therefore infers gender from individuals' names. It uses an algorithm that computes the likelihood that a person with a given name is a woman and classifies profiles where this likelihood is greater than 50 percent as belonging to women. In practice, the predicted female probability is very close to zero or very close to one for a vast majority of all profiles. The imputation of gender is purely name-based and does not consider additional clues such as having a gender-typical education or occupation, thus avoiding potential stereotyping based on such variables.

Age. To impute age, we utilize the date information of the first educational degree in a profile. If the profile indicates a high school degree, we assume the user finished high school at the typical graduation age of 18. On LinkedIn profiles that contain both the year of high school degree and the start years of subsequent studies, a median of zero years elapses between high school graduation and the beginning of undergraduate studies (associate, bachelor), and a median of six years elapses between high school graduation and the start of graduate studies (master, professional, doctorate). For LinkedIn profiles that do not indicate a high school

¹⁴93 percent of the LinkedIn profiles in our estimation sample report an undergraduate degree. In the case of profiles that report only a graduate or postgraduate degree, we compute experience since the start date of the first advanced degree (master, professional degree, doctorate). In cases where the end date of a job is missing, we assume that the employment is ongoing if no subsequent job is listed, or ended with the start date of the following job if a subsequent job is reported.

¹⁵In case a profile reports multiple job spells at the time of the scrape date, we assign the occupation with the highest earnings.

graduation year, we thus assume an age of 18 at the start of the first undergraduate degree, or if that information is missing, an age of 24 at the start of graduate studies.¹⁶

Job-Based and Occupation-Based Earnings. LinkedIn profiles do not report workers' earnings. Therefore, we are unable to observe individual-level earnings that may depend on many idiosyncratic and unobserved factors. However, we use two complementary strategies to infer whether individuals work in jobs or occupations that are typically highly paid.

Our primary such metric is based on Revelio Labs' proprietary salary model that predicts the annual salary for each job from the job title, company name, location, job tenure, and year of observation. This model is trained on salary information in publicly available visa application data, databases of self-reported salary data, and job postings data. It is noteworthy that this earnings imputation is largely based on characteristics of the job (job title, company, location) and considers only tenure as a worker-level earnings determinant. We refer to the earnings data provided by Revelio Labs as 'job-based earnings' to clarify that this measure is informative primarily for between-job but not for within-job earnings differentials across workers. This measure for instance allows us to analyze whether gender differences in skills help to statistically explain the differential representation of women and men across high-paid and low-paid jobs, but we cannot study whether a gender skill gap contributes to gender earnings gaps within a job type.¹⁷

We complement the job-based earnings from Revelio Labs with a measure of occupation-based earnings that we derive from the Census Bureau's American Community Survey (ACS). Using 95 four-digit SOC codes, we compute the average earnings of all employees of that occupation in the pooled 2018 and 2019 ACS data obtained from Ruggles et al. (2024).¹⁸ The occupation-based earnings measure does not exploit the detail of the LinkedIn data as well as the job-based earnings from Revelio Labs, but it provides a useful complementary measure given its simplicity and transparency. Differentials in occupational wage levels have also been previously identified as an important component of the gender wage gap (e.g.,

¹⁶These estimated entry ages by educational level fall within the typical educational entry ages for the US that are reported by the OECD (2021).

¹⁷The construction of Revelio Labs' job-based earnings measure does not consider either worker gender or skills. Therefore, it will assign the same job-typical earnings value to workers of the same job type regardless of their gender or skill sets.

¹⁸In unreported robustness analyses, we alternatively explored occupational earnings metrics that are based only on full-time, full-year employees, or only on male employees. Results for these measures are very similar to those that use earnings from all employees.

Olivetti et al. 2024). Appendix Figure A1 shows that there is a very high correlation between the occupation-based earnings from the ACS and the job-based earnings derived by Revelio Labs (correlation coefficient of 0.89).

2.4.4 Sample Selection and Representativeness

We focus on college graduates in the U.S. because LinkedIn is particularly widely used in that population (Auxier and Anderson 2021). Our estimation sample includes all LinkedIn profiles of U.S.-based workers with a college degree that contain information on skills and other basic variables that are critical for our analysis.

Appendix Table A2 documents the steps of our sampling procedure. In total, we observe 61.9 million profiles of workers who held a job in the U.S. in 2019. As a first step, we drop 6.4 million incomplete profiles for which Revelio Labs was not able to infer baseline data on gender, occupation, or location. We subsequently retain only the three fifths of individuals who indicate valid educational information. Among these, 73 percent report at least one college degree, which confirms the overrepresentation of college graduates on LinkedIn relative to their share of about 50 percent in the overall workforce (Bureau of Labor Statistics 2022). We further restrict the sample to profiles that indicate education dates from which we can infer age, those whose imputed age in 2019 is 23 to 64 years, and those where the start year of the first indicated work experience appears broadly consistent with age and graduation dates.¹⁹ Finally, since the focus of our analysis is on skills, we restrict the analysis sample to the roughly half (49 percent) of the profiles remaining after the previous filtering steps that report at least one skill.

Our final analysis sample contains 8,850,314 profiles. While inclusion in this final sample requires that a profile comprises sufficiently complete information on education, experience, and skills, Appendix Table A2 indicates that the totality of the sample selection steps does not strongly alter the sample composition in terms of age or gender.

To assess the representativeness of the sample in relation to the U.S. college-educated labor force, we compare basic descriptive statistics of our estimation

¹⁹We consider work experience information to be inconsistent with educational timing if imputed age at the start of the first job is lower than 14, or if the duration between graduation from the highest degree and start of the first job is more than five years. While the omitted profiles may contain some valid profiles of individuals with unusual educational and professional career trajectories, we suspect that many of these profiles are either incomplete or contain erroneous data on either education or work years.

sample to representative worker data from the 2019 American Community Survey (ACS). The 8.85 million worker profiles with skills in our sample correspond to 13.6 percent of all U.S. workers with college degrees in the age range of 23 to 64 according to the ACS. The LinkedIn sample is somewhat younger (average age of 37.5 vs. 41.9 years) and more male (54 vs. 48 percent) than the ACS population. Panel A of Appendix Figure A2 shows that the LinkedIn sample also contains considerably more workers with graduate degrees (39 vs. 29 percent). These patterns are consistent with a slightly higher propensity of young people and males to use the platform, and a notable bias of the user base towards more highly educated individuals (Auxier and Anderson 2021).

Occupation groups that tend to be concentrated in the public sector, such as education and health care, are underrepresented in the LinkedIn sample (Appendix Figure A2, Panel B), whereas occupation groups such as management, finance, IT, and media are overrepresented, perhaps because professional networking is seen as particularly important in these fields. The coverage of fields of study (Panel C) indicates a similar pattern with a relative underrepresentation of health and social science majors and an overrepresentation of business and STEM majors. The regional distribution of the LinkedIn observations across the nine U.S. census divisions maps the ACS population quite well (Panel D).

Overall, the LinkedIn sample seems broadly representative of the US workforce with college degrees. While our main results below give equal weight to each worker in the LinkedIn sample, we document in Appendix C that key results remain very similar when we reweight the LinkedIn observations based on population weights for demographic groups taken from the ACS.

2.5 Human Capital Investments and Skills

Classic human capital theory posits that workers acquire skills both through education and through work experience (e.g., Becker 1964; Mincer 1974). If LinkedIn users update their profiles with newly acquired skills over time, we would expect that workers who completed longer educational programs and those with longer periods of work experience report a larger number of skills. In addition, a well-known feature of the Mincer (1974) model is that the relationship between

potential experience and earnings is concave, suggesting that skill accumulation is faster in the years that immediately follow college graduation.²⁰

Another important feature of human capital theory is the distinction between skills accumulated by education and skills acquired through on-the-job experience. This distinction relates directly to the concepts of general versus specific skills: Many types of formal education seek to convey general skills that can be employed in a broad range of occupations, perhaps with the exception of graduate and professional programs that prepare students for particular occupational fields. Work experience instead fosters an accumulation of specific skills that may be specifically relevant for a worker's current occupation. Moreover, work experience is likely also important for the acquisition of managerial skills, which would explain why few workers attain managerial positions right after graduation from college.

Based on these considerations, we formulate the following hypotheses that we test in our cross-sectional data:

1. Workers with higher experience report
 - (a) a larger number of skills, with a concave relationship between experience and skills
 - (b) a larger fraction of specific and managerial skills
2. Workers with more advanced educational degrees report
 - (a) a larger number of skills
 - (b) a larger fraction of specific and managerial skills

We test the first two predictions in section 2.5.1 and the latter two predictions in section 2.5.2.

²⁰See Appendix Figure A3 for an age-earnings profile of US college graduates based on ACS data. The cross-sectional structure of our data prevents us from directly observing the skill accumulation of individual workers over time. Our empirical analysis therefore follows the classical approach of Mincer (1974) who tested his human capital theory by comparing workers with different educational attainment and work experience who were observed at the same time.

2.5.1 Skills by Age and Experience

Number of Skills. One of the most widely documented empirical patterns in labor economics is a concave age-earnings profile that indicates a strong positive relationship between age and earnings at younger ages and a flat or even slightly declining relationship at older ages. In the Mincer (1974) model, this pattern results from on-the-job training that fosters skill acquisition especially at younger ages. If self-reported skills on LinkedIn profiles correspond to the skills considered in this model, we would expect to see a concave age-skill profile.

Intriguingly, the age-skill profile we derive from the LinkedIn data in Figure 2.1 indeed closely mirrors the well-known concave shape of age-earnings profiles. The figure depicts the average number of skills reported by college graduates for two-year age bins ranging from ages 23 to 64. The mean number of skills increases from 14.6 at age 23-24 to 22.3 at age 49-50 and then remains rather flat, declining only weakly at the very end of the age range.²¹

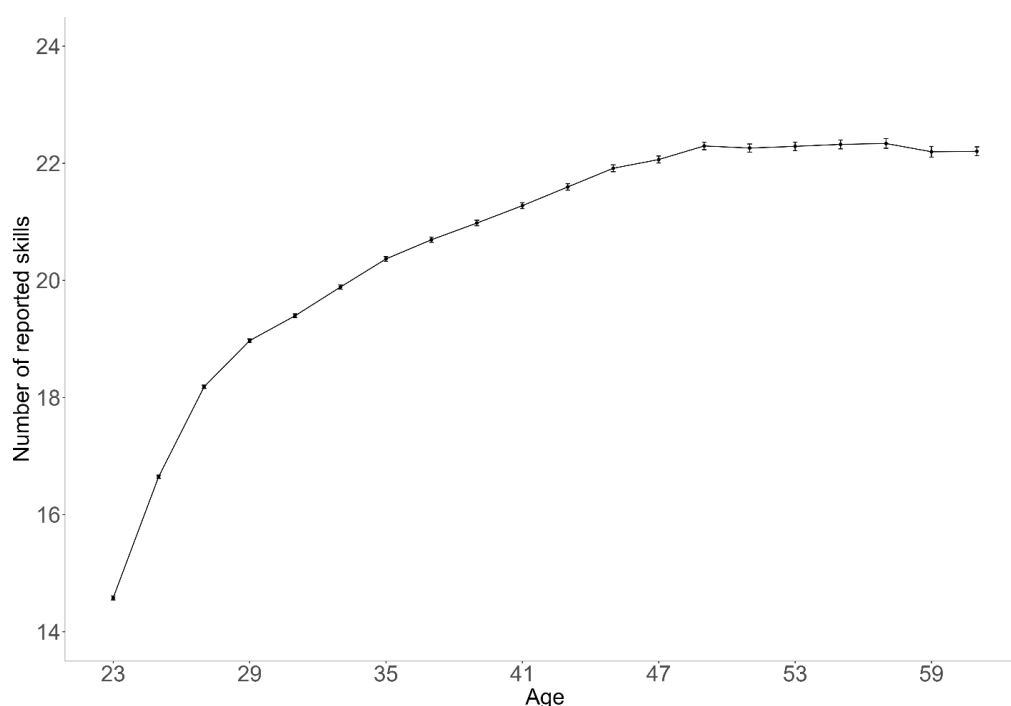
Skill profiles have a similar concave shape when depicted by experience rather than age. Given that all individuals in the sample are college graduates, age is closely aligned with potential experience, i.e., the number of years of work experience that an individual could have accumulated had she been employed without interruption since graduation. Panel A of Figure 2.2 shows that the number of reported skills increases until about 25 years of potential experience and then flattens.

The LinkedIn profiles also allow us to compute years of actual work experience. For a given worker, actual work experience will be lower than potential experience if the worker did not hold a job for every month after graduation from college.²² If skills are acquired primarily through work experience rather than through general life experience as a worker ages, then the number of reported skills should increase more with actual experience than with potential experience. Panel A of Figure 2.2 confirms this prediction. For instance, the average worker with 20 years of actual work experience reports about one skill more than the average worker who has 20 years of potential work experience.

²¹Appendix B demonstrates that a concave shape of the age-skill profile similarly shows in a sample of workers who recently switched their employers, indicating that the concavity is not driven by lack of profile updating among older workers

²²Our measure of actual work experience is likely an upper bound for the true experience value. Since LinkedIn profiles report the months but not exact dates of employment, we can only observe employment breaks that include at least a full calendar month. Moreover, employees may not report transitory absences from a job, such as parental or sickness leave.

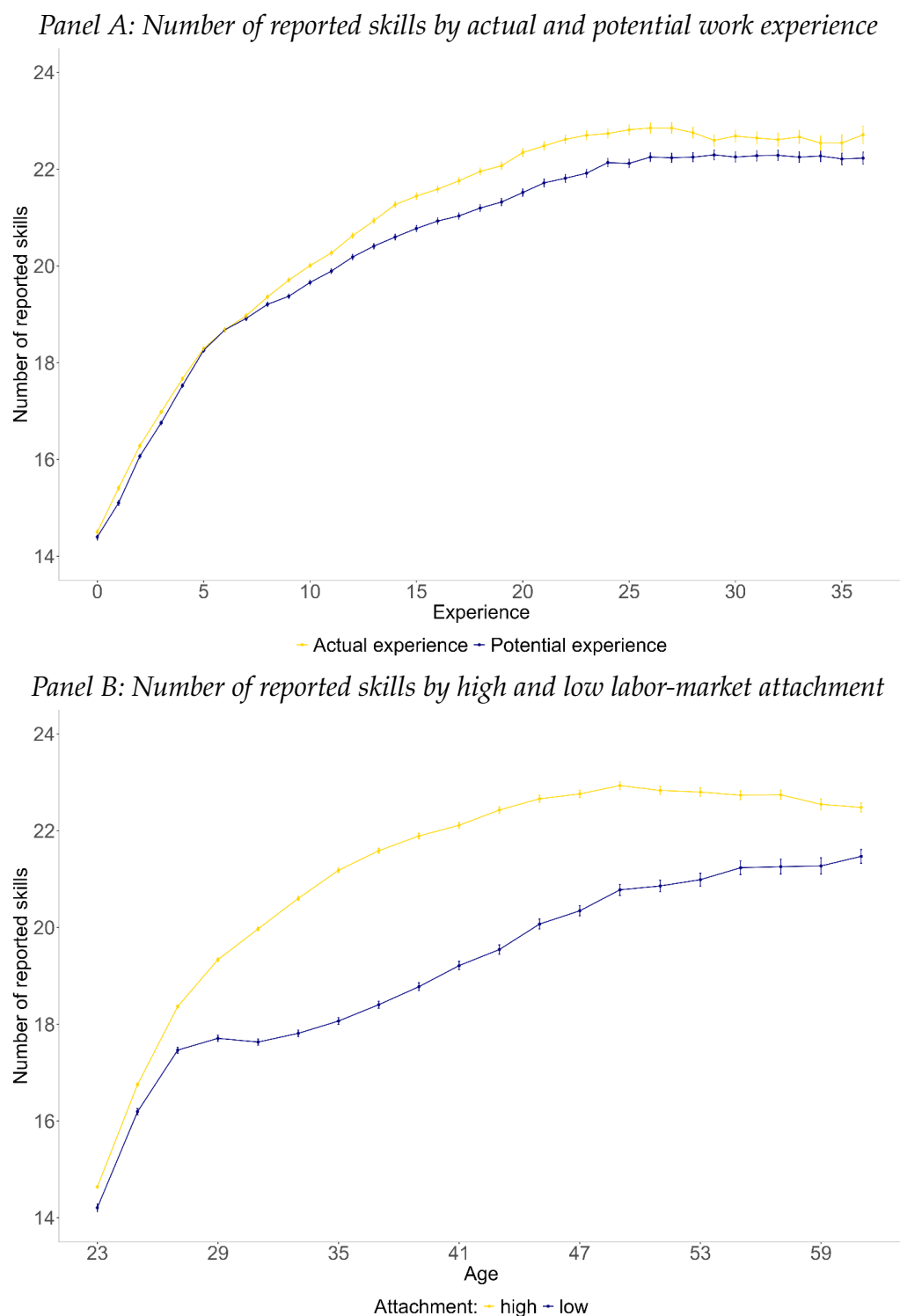
Figure 2.1: Age-skill profile: Number of reported skills by age



Notes: The figure plots the average number of reported skills for two-year age bins, except for the last bin that combines ages 61-64 due to lower sample size. The small error bars depict 99 percent confidence intervals for the skill averages.

As we study workers who were employed in 2019, any difference between actual and potential experience is the result of temporary absences from work between college graduation and the start of the current job. For the median worker in our sample, the ratio of actual to potential experience is 0.93, and this small difference between actual and potential experience helps to explain why the relationships of skills with either potential or actual experience differ only modestly in Panel A of Figure 2.2.

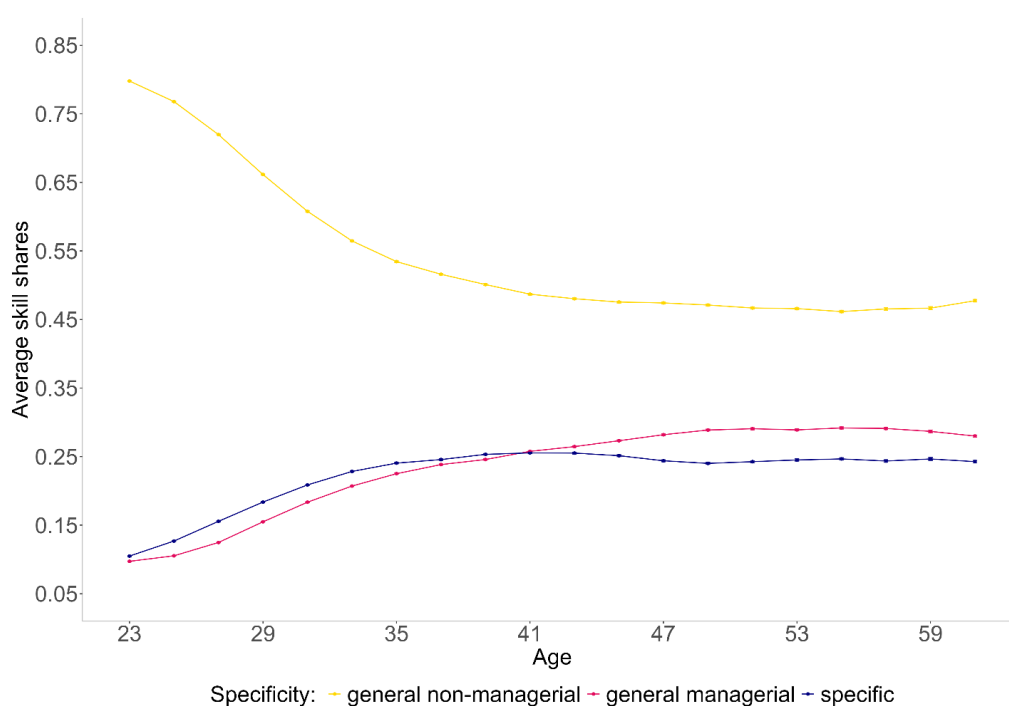
In a complementary analysis, we single out individuals for whom actual experience lags considerably behind potential experience. We classify workers as having a ‘low attachment’ to the labor market if they have a ratio of actual to potential experience below 0.8, and cumulative non-employment spells of at least half a year. Everyone else is classified as high attachment. Panel B of Figure 2.2 shows that for a given age, low attachment workers with significant career breaks report substantially lower levels of skills. The gap is particularly pronounced during the thirties, when the difference between high- and low-attachment workers amounts to over three skills on average.

Figure 2.2: Number of reported skills by experience and labor-market attachment

Notes: Panel A reports the average number of reported skills for workers with the indicated level of potential or actual work experience. Panel B reports the average number of reported skills for workers with low labor force attachment (defined as a ratio of actual to potential experience below 0.8 and cumulative non-employment spells of at least half a year) and for workers with high attachment (everyone else). Error bars depict 99 percent confidence intervals.

Composition of Skills. We next test the hypothesis that younger workers whose human capital was formed primarily through schooling possess primarily general non-managerial skills, while older workers report relatively more specific and managerial skills that may have been acquired on the job. Figure 2.3 confirms these predictions. At age 23-24, 79.8 percent of workers’ skills fall in the general non-managerial category, while this share decreases steadily to 47.5 percent at age 45-46 and remains roughly constant at older ages. The fraction of occupation-specific skills more than doubles from 10.5 percent at age 23-24 to 25.5 percent at age 39-40 and then levels off. The fraction of managerial skills has an even stronger positive age gradient and grows up to a higher age. It triples from 9.7 percent at age 23-24 to 28.9 percent at age 49-50 and remains at similar levels for higher ages.

Figure 2.3: Skill composition by age: Shares of general, managerial, and specific skills



Notes: The figure shows the average fractions of general non-managerial, general managerial, and occupation-specific skills by age. Two-year age bins; last bin combines ages 61-64. Error bars depict 99 percent confidence intervals.

Patterns of skill composition by potential and actual experience and by high and low labor-market attachment again look similar to the patterns by age (Appendix Figure A4). The decline in the fraction of general non-managerial and the increase in specific and managerial skills is stronger for additional years of actual experience than for extra years of potential experience, consistent with

the interpretation that skill sets evolve with on-the-job experience, rather than just with general life experience.

2.5.2 Skills by Education: Degrees, Fields of Study, and College Quality

Classical empirical implementations of human capital models such as Becker and Chiswick (1966) or Mincer (1974) proxied for the amount of human capital obtained from education with years of schooling. Subsequent work argued that human capital may depend not only on education years but also on degrees obtained, fields of study, or attending highly-ranked and prestigious universities (e.g., James et al. 1989; Altonji et al. 2016).

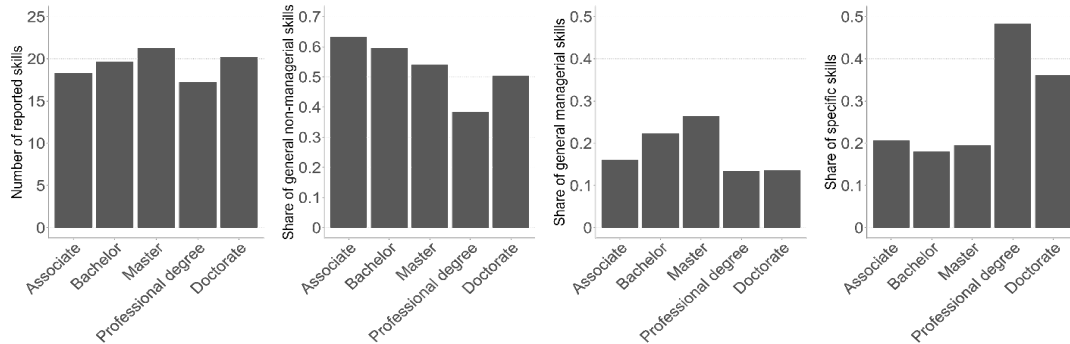
We next conduct descriptive analyses that investigate the number of skills and the shares of general, specific, and managerial skills reported by workers who differ in terms of degree level, field of study, and college ranking. These analyses assess whether greater or different educational investments translate into corresponding patterns in terms of the number and composition of skills.

Panel A of Figure 2.4 indicates that the average number of reported skills is increasing with typical education program length for the three most frequent types of degrees in our data. College graduates with Associate's degrees report an average of 18.3 skills, those with Bachelor's degrees list 19.6 skills, and those with Master's degrees indicate 21.3 skills. If we assume that these graduates' total education consists of 12 years of schooling before entering college, and subsequent studies of two years for an Associate's, four years for a Bachelor's, and six years for a Master's degree, then these numbers imply a nearly linear relationship between education years and skills, with about 0.7 additional skills reported per year of education. However, this linearity is notably violated for the 9 percent of graduates in our sample who hold professional degrees or doctorates and who on average report a lower number of skills than the Master's graduates despite having equal or longer periods of education.²³

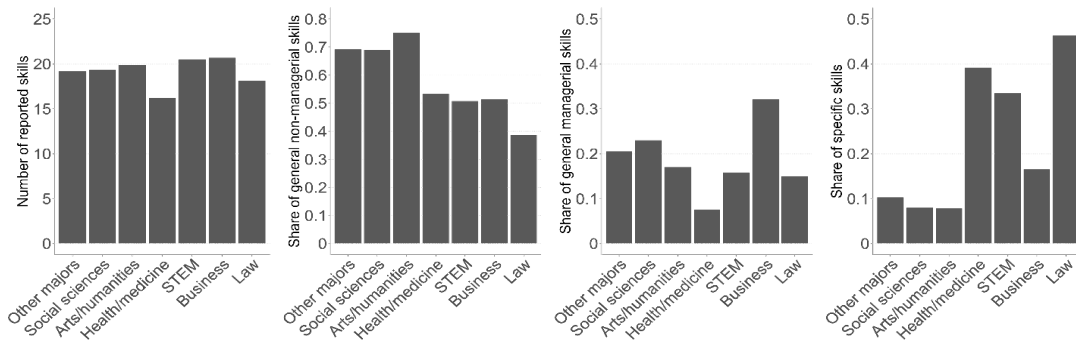
²³A potential explanation for this pattern is that holders of advanced degrees may be less inclined to report basic skills on their profiles, as discussed in section 2.3.3.

Figure 2.4: Number of reported skills and skill composition by education

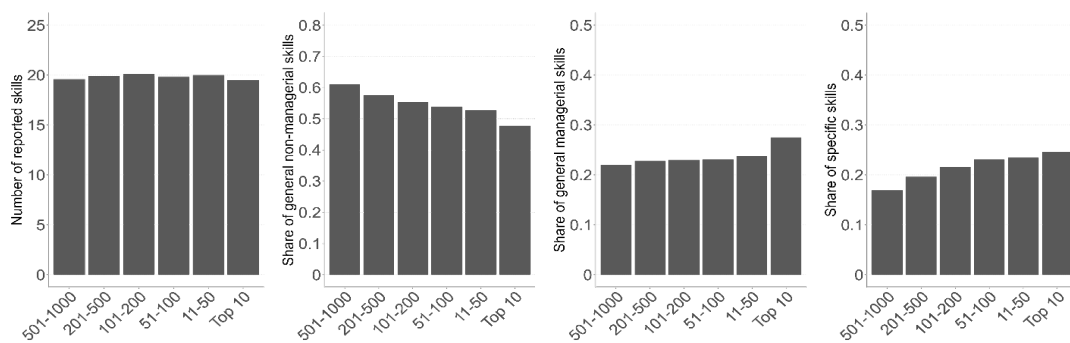
Panel A: Skills by highest educational degree



Panel B: Skills by field of study



Panel C: Skills by college ranking



Notes: Average number of reported skills and average fractions of general non-managerial, general managerial, and occupation-specific skills by highest educational degree, field of study, and Times Higher Education (2019) U.S. college ranking, respectively. Samples: Panel A: full sample; Panel B: sample with field-of-study information (93.1 percent); Panel C: sample of graduates from U.S. colleges listed in Top 1000 of college ranking (68.5 percent).

The holders of professional degrees or PhDs, however, stand out by reporting roughly double the fraction of specific skills compared to those with other degrees. This pattern is consistent with the interpretation that many study programs for

Associate's, Bachelor's, and Master's degrees emphasize general skills that are broadly applicable across a wide range of occupations, while programs for PhD studies or professional degrees such as MDs or MBAs are more targeted towards specific occupational careers and thus convey more specific skills.

We next explore skill patterns for graduates with different fields of study in Panel B of Figure 2.4. For ease of illustration, we aggregate the 37 detailed study subjects observed in the data into seven broader categories which are ordered by ascending average earnings in the figure. While there is only modest variation in terms of the number of reported skills by subject area, there are noteworthy differences in the composition of skills. Individuals who majored in disciplines whose graduates achieve higher average earnings report a higher fraction of specific skills than graduates of fields that are typically associated with lower-paid careers. Specific skills account for 30 percent or more of the skills indicated by individuals with health, STEM, or law degrees. By contrast, fewer than 10 percent of the skills are occupation-specific for graduates of arts and humanities or social sciences. Individuals with a business degree report by far the highest fraction of managerial skills.²⁴

Panel C of Figure 2.4 reports skill statistics by college quality, which we approximate by the U.S. college ranking of Times Higher Education (2019). There is little difference in the total number of skills reported by graduates of differentially ranked institutions. However, the shares of managerial and specific skills in workers' skill sets are considerably higher for those who graduated from top U.S. universities rather than lower-ranked schools.

One challenge for the interpretation of the findings in this section is that they correlate only one dimension of education with skills at a time. It thus is not clear from the evidence of Figure 2.4 whether, for instance, the higher fraction of specific skills among graduates of highly ranked universities reflects a specialization of these universities in study fields or degree programs that are associated with more specific skills, or whether college quality is correlated with a higher share of specific skills also conditional on study fields and types of degrees. In Appendix Table A3, we present multivariate regressions that relate either the total number of reported skills or the shares of general non-managerial, managerial, and specific skills to indicators for degree types, study fields, college ranking, and experience. This

²⁴Although our analysis in the previous section suggests that managerial skills—like specific skills—are often acquired through on-the-job experience, there is evidence that business school students obtain greater managerial skills already in college (Kang and Sharma 2012).

exercise suggests that the results presented in this and in the previous section also hold as conditional correlations: While there is relatively modest variation in the number of skills reported by college graduates whose education differs in terms of degrees, field of study, or college ranking, there are notable differences in the composition of skills. Holders of advanced degrees, those studying science and professional fields, and graduates of highly-ranked universities report higher shares of occupation-specific skills, and in some cases, higher shares of managerial skills. Greater actual work experience is associated with a higher number of skills and larger shares of specific and managerial skills.

2.6 Skills and Job-based Earnings

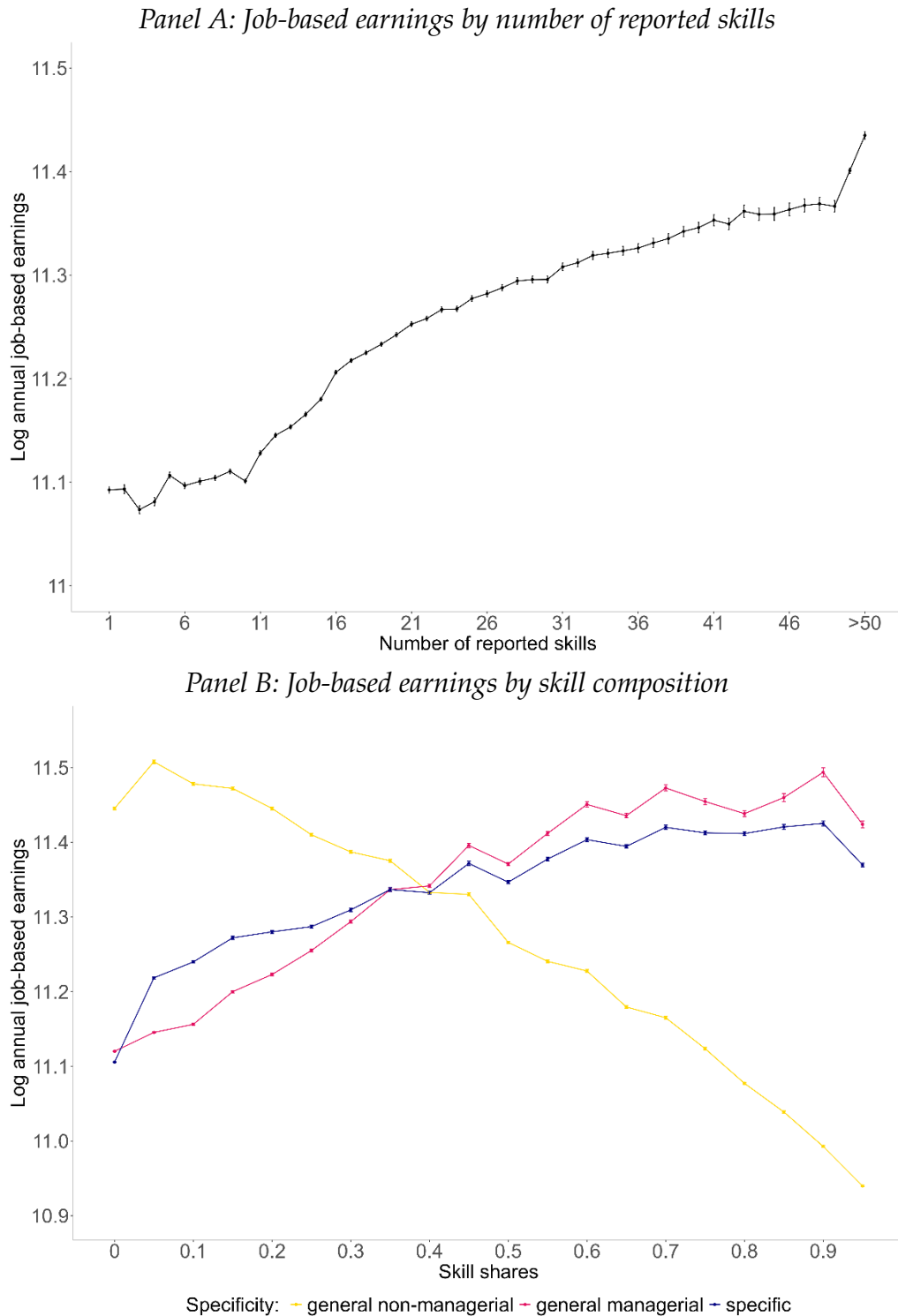
After having established that skills are systematically related to human capital investments from which these skills presumably result, we investigate the second foundational tenet of human capital theory: individuals with greater skills should command higher earnings. Section 2.6.1 studies whether workers with more skills, and those with higher fractions of specific and managerial skills, are employed in higher-paid jobs. We then analyze which detailed skills are most strongly associated with higher-paid jobs (section 2.6.2) and assess the extent to which differences in skill sets help to account for the variation in job-based earnings across workers (section 2.6.3).

2.6.1 Number of Skills, Skill Composition, and Earnings

The simplest metric of the skill sets reported on LinkedIn profiles is the number of reported skills. The results in Figure 2.1 on the average number of reported skills by age display a remarkable similarity to the well-documented concave shape of age-earnings profiles. We thus begin our analysis of the relationship between skills and earnings by investigating whether individuals who report a larger number of skills tend to have higher job-based earnings.

Results show that workers who report more skills are indeed more likely to have higher-paid jobs. Panel A of Figure 2.5 indicates that log earnings are consistently and, over a broad range of skill levels, nearly linearly increasing in the number of reported skills. On average, each additional skill is associated with a 0.67 log point increase in annual job-based earnings.

Figure 2.5: Job-based earnings by number of reported skills and skill composition: Unweighted vs. weighted data



Notes: Average log annual job-based earnings are imputed by Revelio Labs’ proprietary salary model. Panel A indicates average earnings for all LinkedIn profiles with the indicated number of skills, while Panel B indicates average earnings for five-percent bins of general non-managerial, general managerial, and occupation-specific skills. Error bars depict 99 percent confidence intervals.

We established in section 2.5 that workers with more advanced educational degrees and those with more work experience report larger fractions of specific and managerial skills and a lower share of general non-managerial skills. Panel B of Figure 2.5 indicates that job-based earnings vary strongly with the skill mix across these broad skill domains. Individuals who report a larger share of occupation-specific skills tend to have substantially higher imputed earnings. Similarly, and even more strongly, earnings increase in the share of managerial skills. Consequently, the earnings gradient is negative in the share of general non-managerial skills.

The patterns for the association between skill categories and earnings in Panel B of Figure 2.5 help us to qualify the section 2.5.2 results on the relationship between education and skills. While we observed little variation in the number of reported skills for different fields of study, we now note that the primarily general skills reported by graduates in humanities and social sciences appear to be less valued in the labor market than the more specific skills of science, medicine, and law graduates or the managerial skills of business graduates.

2.6.2 Multidimensional Skills and Earnings

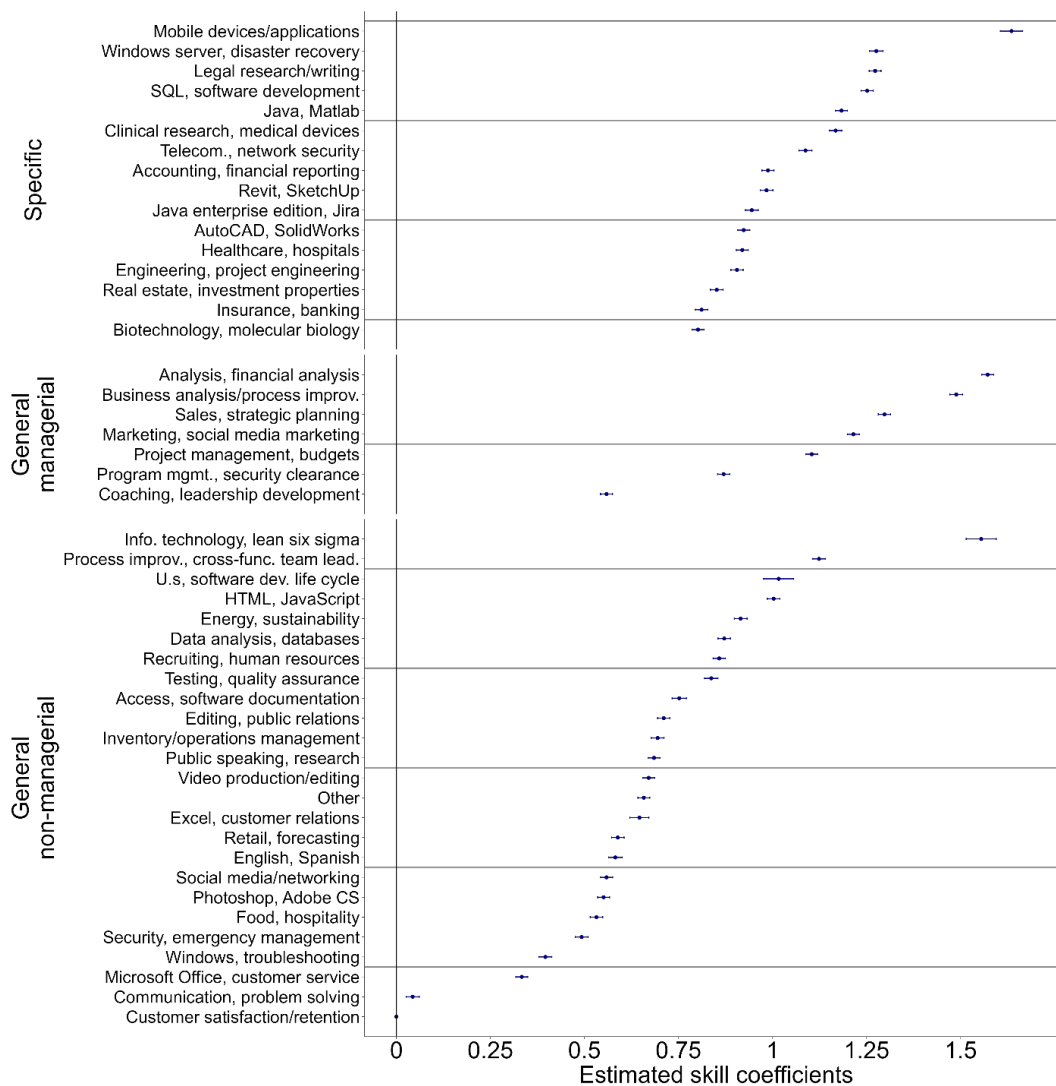
We more formally assess the association of different types of skills with job-based earnings by regressing workers' imputed log earnings on the total number of reported skills and the shares of skills reported in each of the detailed skill clusters listed in Table 2.1.²⁵ Figure 2.6 shows the estimated regression coefficients for each skill cluster, ordered by decreasing coefficient size within the three broad skill categories (specific, managerial, general non-managerial). The omitted reference category for the skill share variables is the cluster 'customer satisfaction/retention', which has the weakest positive association with job-based earnings.

The results in Figure 2.6 corroborate the previous observation that larger shares of specific and managerial skills are associated with higher job-based earnings than general skills. About four-fifths of the clusters of specific and managerial skills have above-average earnings coefficients. Workers with specialized IT skills ('mobile devices, mobile applications', 'Windows server, disaster recovery', 'SQL, software development', 'Java, Matlab'), legal skills ('legal research, legal writing'), medical

²⁵We obtain qualitatively similar results from a regression that relates log earnings to the number of skills reported in each of the 48 skill clusters. The model with total number of skills and shares of skills by cluster however achieves a better goodness of fit as it is less sensitive to outlier observations from individuals who report an unusually large number of skills in a given cluster.

skills (‘clinical research, medical devices’), and various types of managerial skills (‘analysis, financial analysis’, ‘business analysis, business process improvement’, ‘sales, strategic planning’, ‘marketing, social media marketing’) obtain substantially higher earnings than other workers, with each one percentage-point increase in the share of skills in these clusters being associated with a 1.1 to 1.6 log point greater earnings premium relative to skills in the baseline skill category of ‘customer satisfaction/retention’.

Figure 2.6: Association between job-based earnings and skills



Notes: The figure plots estimated skill coefficients and 99-percent confidence intervals from an individual-level regression of log annual job-based earnings (in log points) on the total number of reported skills and shares of these skills pertaining to each of 47 skill clusters (in percentage points). The omitted skill cluster share in the regression is ‘customer satisfaction/retention’. See column 4 of Appendix Table A4 for tabulated coefficients.

Conversely, four-fifths of the general skill clusters have below-average coefficients in the earnings regression. Estimated earnings coefficients are smallest for the skill clusters ‘customer satisfaction, customer retention’ (the reference category), ‘communication, problem solving’, ‘Microsoft Office, customer service’, and ‘Windows, troubleshooting’, which comprise relatively basic customer interaction, communication, and standard software skills. The few general skill clusters that have large earnings coefficients instead comprise more advanced technology and process management skills (‘information technology, lean six sigma’, ‘process improvement, cross-functional team leadership’) that are used across a broad range of occupations.

2.6.3 Do Skills Explain More Earnings Variation than Standard Human Capital Measures?

The systematic relationships between skills reported on LinkedIn profiles and job-based earnings strengthen the interpretation of the self-reported skills as a metric for human capital that is valued in the labor market. But are these skill measures better at explaining earnings variation than more conventional proxies for human capital such as experience or education? This should be the case if employers reward workers’ skills rather than their past investments into skill acquisition, and if high-quality measures of skills, experience, and education are available. However, to the extent that skills on LinkedIn profiles are reported noisily as discussed in section 2.3.3, it is conceivable that measures of skill investments continue to provide better proxies for the true level of worker skills and thus explain more earnings variation.

We study earnings regressions that relate job- or occupation-based earnings to either a set of education and experience variables in the spirit of Mincer (1974), or the LinkedIn skill measures, or both. These regressions ask whether workers with specific individual human capital attributes (e.g., higher levels of education, experience, or skills) are overrepresented in highly paid jobs or occupations. We exclude job-level predictors of earnings such as job titles or occupations from these regressions for two reasons. First, our conceptual aim is to assess the contribution of worker-level human capital to earnings variation in the labor market, rather than to study the role of job characteristics. Second, the outcome variables for job-based or occupation-based earnings are derived from job attributes, so including such predictors would introduce mechanical correlation with the dependent variables.²⁶

²⁶While we therefore do not study the contribution of occupations to variation in job-based earnings, we can assess the explanatory power of human capital variables for within-occupation

We begin in in Panel A of Table 2.2 with a basic OLS regression model in the spirit of the classical Mincer equation that relates log earnings to years of education and a quadratic of years of potential experience. Since our sample consists entirely of college graduates, we replace the linear term for education years with a set of indicator variables for different types of college degrees that are typically associated with shorter or longer durations of study (Associate, Bachelor, Master, Professional, and PhD degrees). This basic Mincer model explains 14.2 percent of the overall variation in job-based earnings in our sample.²⁷ The second regression model in Table 2.2 relates log earnings to a much richer set of education and experience measures by adding 26 indicators for detailed fields of study, eight indicators for the placement of the degree-granting institution in the Times Higher Education U.S. college ranking, and a quadratic in workers' actual (in addition to potential) work experience. These measures of education and experience are unusually detailed. While we can observe them in the LinkedIn data, many of these variables would not be available in individual-level datasets that are more widely used for labor-market analyses, such as the Current Population Survey, the American Community Survey, or the Longitudinal Employer-Household Dynamics data. The augmented specification, which includes a total of 42 education and experience variables, explains 23.5 percent of the variation in job-based earnings. In the third column of Table 2.2, we explore whether the explanatory power can be increased further by using random forests to flexibly model the association of earnings with the large set of education and experience variables.²⁸ To generate comparable R^2 statistics for both OLS and random forests, all regression models in Table 2.2 are based on a randomly selected 70% of all observations for which we fit the OLS and random forest models, and we compute goodness-of-fit statistics out-of-sample by assessing the predictive power of model-predicted values in the 30% holdout

variation in job-based earnings. To do so, Panel C of Table 2.2 uses the residuals from a regression of the Revelio Labs earnings measure on four-digit occupation indicators as the outcome variable.

²⁷To conserve space, Table 2.2 only indicates goodness-of-fit statistics for the different regression models, while Appendix Table A4 reports the full set of coefficient estimates for the OLS models.

²⁸We use standard random forests as introduced by Breiman (2001) and estimated via the ranger package in R (Wright and Ziegler 2017) to model the association between earnings and varying sets of control variables in Table 2.2. Random forests are ensemble methods that combine many decision trees to improve predictive accuracy and reduce overfitting (Hastie et al. 2009). We use 500 trees to build each forest to trade off model performance and computational efficiency. Comparing results using different numbers of trees, we found that R^2 estimates stabilize from around 100 trees upward.

sample.²⁹ The column 3 result however shows no additional gain in explanatory power when using random forests instead of OLS to model the relationship of earnings with detailed education and experience variables.³⁰

The regression models in the fourth to sixth columns of Table 2.2 replace the education and experience variables that measure inputs into human capital production with the skill variables observed on LinkedIn that measure an output of this production. Column 4 uses a parsimonious specification that relates log earnings to the number of reported skills and the shares of specific and managerial skills among total skills. This basic model with just three skill variables accounts for more variation in job-based earnings than the basic Mincer specification (17.2 vs. 14.2 percent). The more detailed skill model in column 5 considers the share of skills that a worker reports across the 48 skill clusters (corresponding to the OLS regression underlying Figure 2.6), while column 6 more flexibly models the association between earnings and detailed skill variables using random forests. The corresponding estimates indicate that the models with skill variables outperform those with detailed education and experience variables in terms of explanatory power. With random forest specifications, the fraction of job-based earnings variation explained by skills is one-fourth larger than the fraction explained by education and experience (29.7 vs. 23.4 percent).

The final two columns of Table 2.2 relate earnings to the full set of education, experience, and skill variables using either OLS or random forest estimation. Relative to the random forest model with detailed education and experience variables in columns 4, the addition of skill variables in column 8 raises the explanatory power by more than one-half from 23.4 percent to 36.3 percent. This sizable improvement suggests that the skill variables provide substantial additional information about workers' human capital that cannot already be inferred from detailed metrics of education and experience.

By contrast, the gain from adding education and experience variables to a skills-only model is considerably smaller (6.6 percentage points for the random

²⁹While it is also possible to compute out-of-bag pseudo- R^2 values using random forests, these are not directly comparable to the adjusted R^2 statistics from OLS. Pseudo- R^2 reflects predictive performance based on internal cross-validation within trees, whereas OLS adjusted R^2 penalizes for model complexity. In our application, the reported out-of-sample R^2 statistics differ only minimally from out-of-bag pseudo- R^2 . The R^2 values for the OLS models also barely differ between the split sample approach used in Table 2.2 and standard regressions with the full sample of observations.

³⁰The random forest model has a better goodness-of-fit than OLS in the 70% training sample, but a marginally weaker fit in the out-of-sample evaluation for which the R^2 values are reported in Table 2.2.

Table 2.2: Explaining earnings variation with alternative metrics of human capital

	Education/experience			Skills			Combined	
	Basic OLS (1)	Detailed OLS (2)	Random forest (3)	Basic OLS (4)	Detailed OLS (5)	Random forest (6)	Detailed OLS (7)	Random forest (8)
<i>Regressors</i>								
Potential experience (squared)	✓	✓	✓				✓	✓
Actual experience (squared)		✓	✓				✓	✓
Highest degree	✓	✓	✓				✓	✓
Field of study		✓	✓				✓	✓
College ranking		✓	✓				✓	✓
Number of skills				✓	✓	✓	✓	✓
Shares of specific and managerial skills				✓	(✓)	(✓)	(✓)	(✓)
Shares of skills by skill clusters					✓	✓	✓	✓
A. Job-based earnings (Revelio Labs)								
Adjusted R^2	0.142	0.235	0.234	0.172	0.253	0.297	0.325	0.363
B. Occupation-based earnings (ACS)								
Adjusted R^2	0.097	0.192	0.196	0.113	0.230	0.267	0.276	0.319
C. Within-occupation variation in job-based earnings (Revelio Labs)								
Adjusted R^2	0.078	0.109	0.099	0.071	0.106	0.130	0.147	0.172

Notes: $N = 8,850,314$ LinkedIn profiles randomly split into a 70% training and 30% holdout sample. The table reports the R^2 goodness-of-fit statistic (adjusted for number of regressors) for regression models whose dependent variable is either log annual earnings imputed by Revelio Labs' proprietary salary model ('job-based earnings'), the average log annual earnings at the SOC four-digit occupation level based on 2018-2019 American Community Survey (ACS) data ('occupation-based earnings'), or the residuals of a regression of job-based earnings on four-digit occupation fixed effects ('within-occupation variation in job-based earnings'). To ascertain comparability in goodness-of-fit statistics across OLS (columns 1-2, 4-5, 7) and random forest methodologies (columns 3, 6, 8), all models are estimated on a 70% training sample, and goodness-of-fit is determined using model-predicted values in the 30% holdout sample. Regressors are defined as follows: Potential experience (squared) is the time between college graduation and scrape month (in years) and its square. Actual experience (squared) is the cumulated time of job spells in LinkedIn profile (in years) and its square. Highest degree controls for four indicator variables of educational degree levels. Field of study controls for 26 indicator variables based on the two-digit CIP taxonomy. College ranking controls for eight indicator variables based on the Times Higher Education ranking. Number of skills controls for the count of skills reported on the LinkedIn profile. The column 4 model controls for the share of specific skills and the share of managerial skills, while the models in the subsequent columns control in more detail for the share of skills in 47 skill clusters (with the share of skills in the 48th cluster being the omitted category). Regressions include fixed effects for scrape months. See Appendix Table A4 for a tabulation of coefficient estimates from OLS models for job-based earnings estimated on the full sample of LinkedIn profiles.

forest models of columns 6 and 8, or 7.2 percentage points for the OLS models of columns 5 and 7). This observation implies that the impact of education and experience on earnings as measured in columns 2 and 3 is to a substantial extent mediated through skills. A closer inspection of OLS coefficient estimates, which are tabulated in Appendix Table A4, indicates that notably the field-of-study variables lose explanatory power in models that also account for skills. In the column 2 model that considers only education and experience, the three study fields with the largest earnings effects (computer science, engineering, and business studies) are on average associated with 36.4 log points higher earnings than the three lowest-paying fields (education, family and consumer sciences, and public administration). This difference shrinks by two-thirds to just 12.6 log points in the model that additionally controls for worker skills.³¹ Conversely, the coefficient difference between the three skill clusters with highest and lowest earnings associations (1.46 log points in the skills-only model of column 5) declines by less than a quarter in the model that also accounts for education and experience (1.13 log points in column 7).³²

Panel B of Table 2.2 repeats the same set of earnings regressions for the alternative earnings outcome measure that is based on average occupational earnings in the Census Bureau's 2018-2019 ACS data, instead of Revelio Labs' earnings model of job-based earnings. The pattern of the goodness-of-fit statistics across the column 1 to 8 models remains qualitatively similar to the results with the job-based earnings metric in the upper panel of the table. The main quantitative difference is that human capital variables explain a lower fraction of variation in occupation-based earnings than in job-based earnings. This finding implies that the education, experience, and skill variables considered help to explain earnings variation within occupations, for instance across jobs with or without supervisory functions, or those in different firms and locations. Panel C of Table 2.2 confirms this conjecture by relating within-occupation variation in earnings, obtained as residuals from a regression of the

³¹In the model without skill variables, the average earnings coefficient of the three high-paying fields is 28.0 log points whereas the average coefficient for the three low-paying fields is -8.3 log points (relative to the reference category of agricultural sciences). In the combined model that includes skill variables, the average earnings coefficients are 8.2 log points for the former and -4.4 log points for the latter fields.

³²As indicated in Figure 2.6, the three skill clusters with highest earnings coefficients are 'mobile devices/applications', 'analysis, financial analysis', and 'information technology, lean six sigma'. The average coefficient estimate for these skills is 1.59 in the skills-only model and 1.24 in the combined model. The three clusters with lowest earnings coefficients are 'customer service/retention', 'communication, problem solving', and 'Microsoft Office, customer service', with an average coefficient of 0.12 in the skills-only model and 0.11 in the combined model.

Revelio Labs job-based earnings measure on fixed effect for four-digit occupations, to the different sets of human capital variables. While the contribution of these variables to within-occupation variation in job-based earnings is modest overall, the models with detailed skill variables again explain an about equally large or greater share of the earnings differences than detailed education and experience variables.

Our overall conclusion from the analyses in sections 5 and 6 is that skills observed on LinkedIn profiles are related in plausible ways to human capital investments such as education and experience, but contain additional information on workers' human capital that cannot be directly inferred from the education and experience variables alone. As such, worker skills provide a useful complement to education and experience when measuring human capital, and may provide a superior metric if only one type of variables were available.

2.7 The Gender Skill Gap

We conclude our analysis by studying skill patterns by gender. Extant evidence indicates that women and men differ only modestly in terms of traditional human capital proxies such as years of education and experience, while there is an important gender gap in occupational choice that explains a substantial part of the gender earnings gap (Blau and Kahn 2017; Olivetti et al. 2024). Our results of the previous sections suggest that the LinkedIn skill data can capture additional human capital differences that are not captured by education and experience variables, and that these differences in skills help to explain the employment of workers in higher- versus lower-paid jobs and occupations. We bring these insights to bear on the analysis of gender-specific patterns in the labor market by first documenting and discussing gender differences in reported skills (section 2.7.1) and then analyzing to what extent these skill differences contribute to a gender gap in job- and occupation-based earnings (section 2.7.2).

2.7.1 Skills by Gender

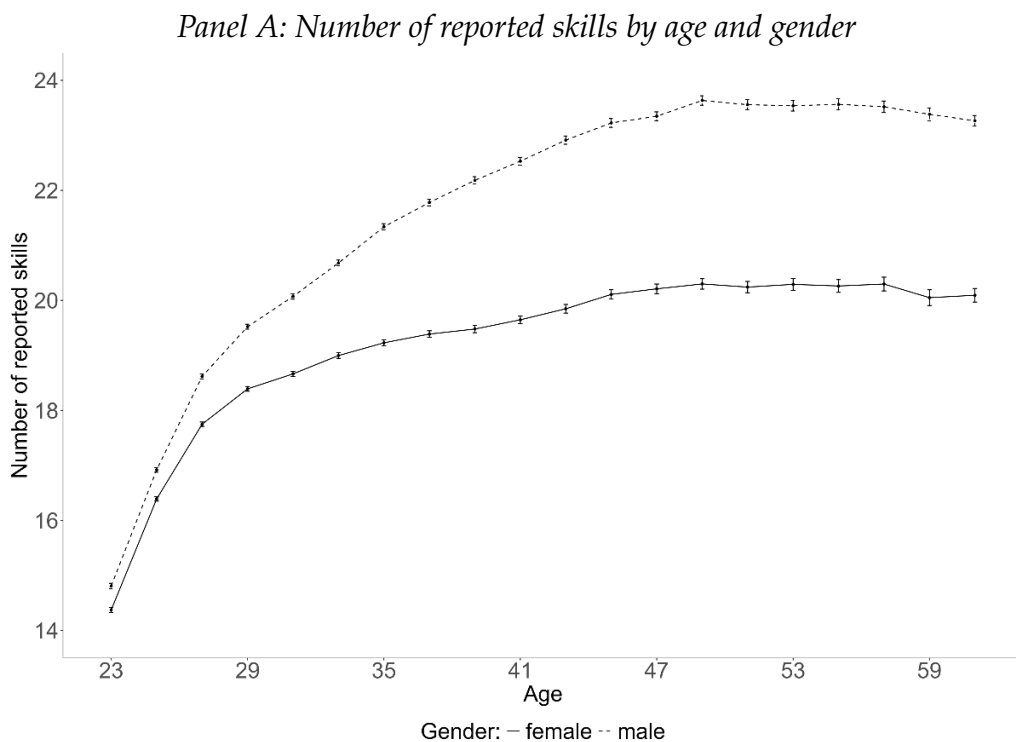
We showed in Figure 2.1 that the cross-sectional relationship between age and the average number of reported skills has a remarkable similarity to the well-documented concave shape of age-earnings profiles. Panel A of Figure 2.7 repeats this analysis separately for women and men. It documents a striking new pattern of gender differences in the labor market: Women report fewer skills than men,

especially at higher ages. The slopes of the gender-specific age-skill profiles are quite similar for both genders in the mid-twenties when these profiles slope steeply upwards, and from the mid-forties, when the profiles flatten and ultimately decline slightly. There is a remarkable difference, however, between the late twenties and early forties where the age-skill profile increases much less for women than for men. While the gender difference in the average number of reported skills is 0.5 in the age range 23-26, it rises steadily to 3.1 at age 43-44 and remains in the range 3.1 to 3.3 until age 64. These gender-specific age-skill profiles are similar to the shapes of gender-specific age-earnings profiles (see Appendix Figure A5).

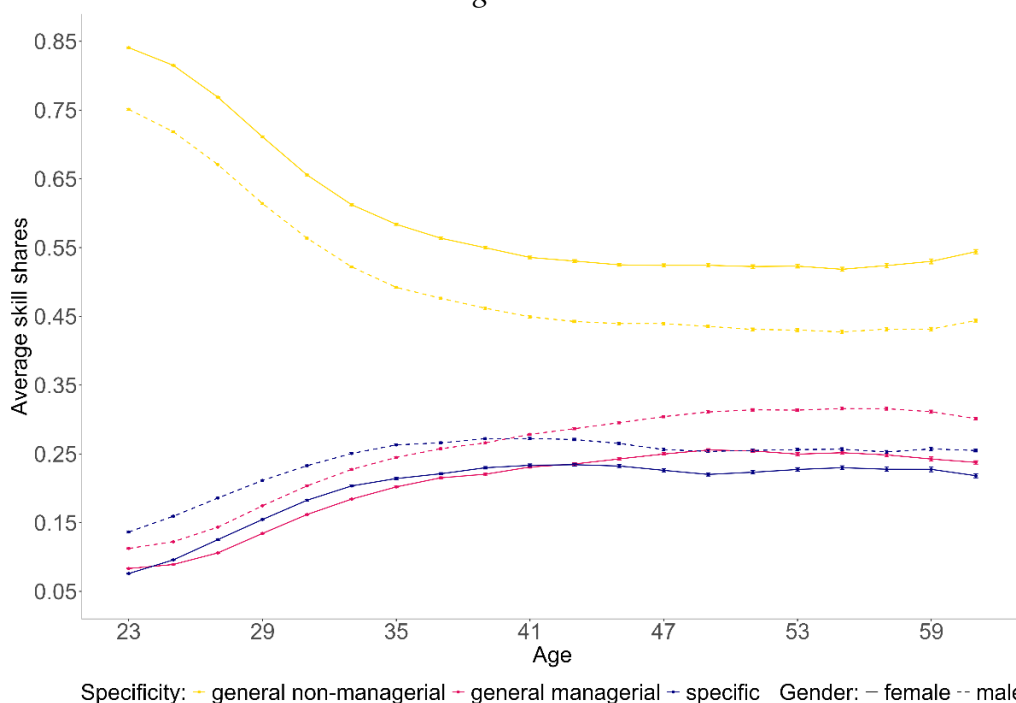
Panel B of Figure 2.7 indicates the shares of general, managerial, and specific skills reported by women and men of different ages. Women consistently report a higher fraction of general skills and lower shares of managerial and specific skills than men of the same age. This pattern suggests that women enter the labor market with a considerably different skill mix than men, which may in part be explained by gendered patterns of field-of-study choice while in college. As shown in Figure 2.4, individuals who majored in the relatively female-dominated fields of social sciences, arts, and humanities report higher shares of general skills than those who majored in STEM, business, health, or law. The difference in the share of specific skills narrows slightly with age, whereas the difference in the share of managerial skills slightly widens.

We next discuss four potential explanations for the observed gender skill gap: male overconfidence, lack of profile updating among women, occupation-specific skill accumulation, and lower skill growth due to reduced labor supply induced by motherhood.

Figure 2.7: Number of reported skills and skill composition by age and gender.



Panel B: Share of general non-managerial, general managerial, and specific skills by age and gender



Notes: The figures show the average number of reported skills and the average fractions of general non-managerial, general managerial, and occupation-specific skills, respectively, by age and gender. Two-year age bins; last bin combines ages 61-64. Error bars depict 99 percent confidence intervals.

Male Overreporting. Experimental evidence indicates that men are overconfident in their abilities (Niederle and Vesterlund 2007) and more likely to self-promote (Exley and Kessler 2022). There is ongoing debate on where and how this phenomenon is quantitatively important, as gender differences in overconfidence may be sensitive to the subject domain (Bordalo et al. 2019) and close to zero on average (Bandiera et al. 2022). Since there is no objective benchmark for the self-reported multidimensional skills that we analyze, we cannot directly quantify the contribution of overconfidence and self-promotion, if any, to the gender gaps in self-reported skills of Figure 2.7. However, two results speak against a predominant role of male overreporting. First, if gender differences in the number of reported skills were driven primarily by male overreporting, it would have to be the case that such overconfidence is much more prevalent among older individuals but nearly absent among younger ones. We are not aware of any literature that establishes such a strong age or cohort gradient in gender-specific overconfidence.

Second, we test whether skills reported by men are as strongly linked to higher-paying jobs as those reported by women. If men were widely overreporting, employers—who gather more accurate skill information during recruitment and on the job—would discount men’s self-reported skills. This would result in smaller earnings gains per reported skill for men compared to women. However, gender-specific earnings regressions (using the specification of column 4 in Appendix Table A4) show similar associations between reported skills and pay, with a slightly higher coefficient for men (0.402) than for women (0.374). Thus, employers do not seem to discount the skills reported by men.

Gender Differences in Profile Updating. A different type of reporting bias emerges when women update their LinkedIn profiles less frequently than men. In this case, the average man’s profile may indicate the current skill set while the average woman’s profile shows the skills that she possessed a few years ago, thus failing to account for the skill accumulation that occurred since. If one assumes that both men and women are especially likely to update their profiles when they are looking for a new job, then the profiles of both genders should be equally up-to-date at young ages where all college graduates had to look for a job, while males’ profiles may be more current at later ages due to greater subsequent male job mobility (Loprest 1992; Keith and McWilliams 1999; Manning and Swaffield 2008) or stronger male career aspirations (Azmat et al. 2025).

In Panel A of Figure 2.8, we analyze a subsample of workers who report an employer change in the last two years, and for whom we thus know that their profiles have been updated relatively recently. The recent firm switchers have steeper age-skill profiles, which could be consistent both with a more thorough updating of the skill information on their LinkedIn profiles, or a positive selection effect where workers who rapidly acquire new skills are more likely to seek better jobs at new employers as job mobility is an important source of earnings growth (Topel and Ward 1992; Guvenen et al. 2021). The focus on the subsample of recent firm switchers however does not narrow the gender skill gap which remains similar in timing and magnitude as in the full sample of LinkedIn profiles. We therefore conjecture that gender-specific patterns of profile updating are not the main driver of the gender skill gap seen in Panel A of Figure 2.7.³³

Between-Occupation Differences in Skill Accumulation. Occupations may vary in the extent that they provide workers with a potential for on-the-job skill accumulation. The flatter slope of women’s age-skill profile may thus result from a concentration of women in occupations that offer less learning opportunities and thus lower returns to experience over the life cycle (Adda et al. 2017).³⁴

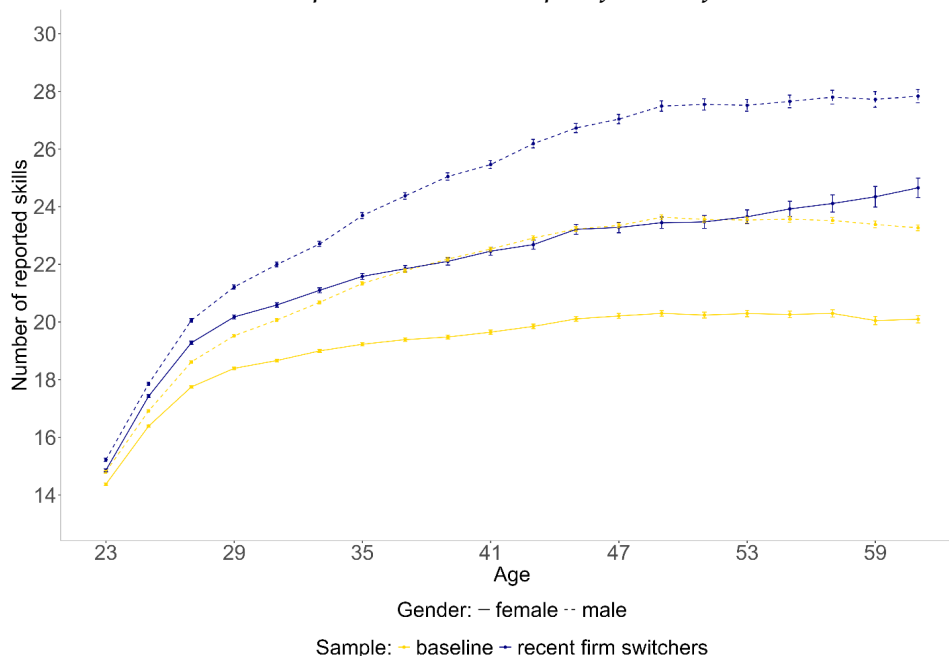
To assess this hypothesis, Panel B of Figure 2.8 indicates average skills by age group for women and men, as in Panel A of Figure 2.7, but additionally shows a data series that weights women’s average reported number of skills by occupation and age with the occupational employment shares of men in the corresponding age group. If the gender skill gap resulted entirely from women and men’s differential distribution across occupations, the counterfactual data series (‘female weighted by male occupational composition’) should match the data series for males. The results show that up to age 28, the small gender gap in reported skills is indeed entirely a between-occupation gap, while women and men in the same occupation report the same average number of skills. However, after age 28, a sizable within-occupation gender gap in skills begins to accumulate until about age 40, and then stays constant. Consequently, gender differences in occupational choice can account for only a small fraction of the overall gender gap in reported skills at older ages.

³³Appendix B provides additional statistics and results for the subsample of recent firm switchers.

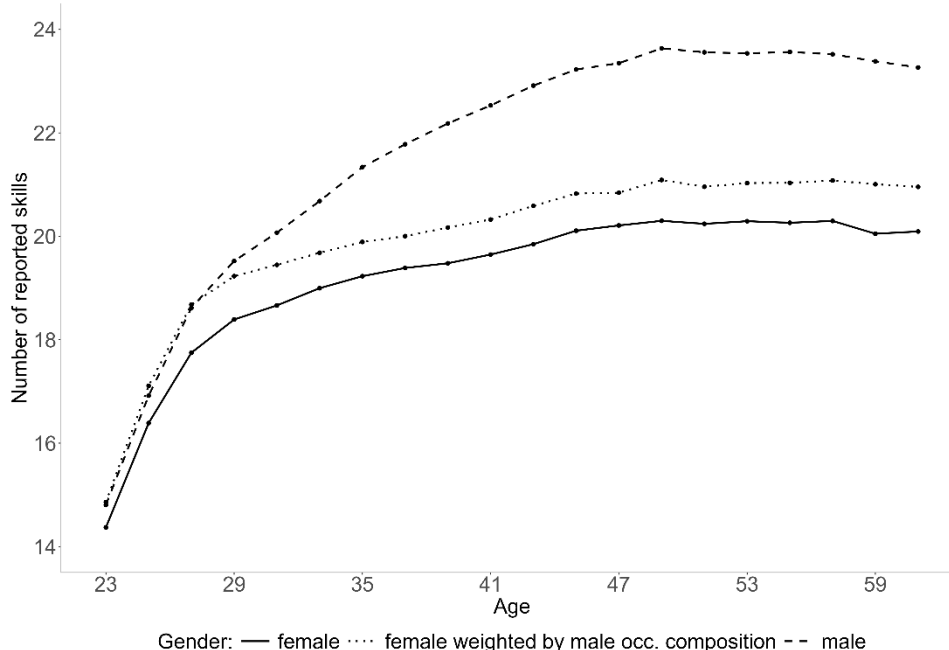
³⁴Adda et al. (2017) show that women are more likely to choose occupations with lower returns to experience at labor market entry in anticipation of later motherhood. There is also evidence that women move to more family-friendly occupations later in their career after childbirth (Kleven et al. 2019), which intensifies occupational sorting by gender and perhaps concentrates women in occupations whose workers tend to report lower numbers of skills.

Figure 2.8: Alternative age-skill profiles by gender

Panel A: Full sample versus subsample of recent firm switchers



Panel B: Correcting for gender differences in occupational composition



Notes: The figures show the average number of reported skills by age and gender. In Panel A, the data series ‘recent firm switchers’ retains only the LinkedIn profiles of individuals who report a change of employer within the last two years. In Panel B, the data series ‘female weighted by male occ. composition’ is based on the average number of skills for women of a given age group and detailed occupation, weighted by the employment share of each occupation among the men of the same age group. The difference between this data series and the male series indicates within-occupation differences in skills across gender, while the difference with regard to the female series indicates between-occupation differences in skills across gender. Two-year age bins; last bin combines ages 61-64.

Appendix Figure A6 shows that a large part (though not all) of the gender gaps in shares of general, managerial, and specific skills at young ages are also between occupations, while within-occupation differences become more important at higher ages. We thus conclude that whereas female and male workers in the same occupation report very similar skill sets at young ages, women subsequently appear to experience less within-occupation growth in overall, managerial, and specific skills.

Gender Differences in Labor Supply and Effects of Parenthood. An important gender difference in career trajectories is that women tend to have lower work hours and more career interruptions than men, especially after becoming parents (Bertrand et al. 2010; Cortés and Pan 2023). Following childbirth, mothers' earnings decline substantially relative to fathers' earnings, and this 'motherhood penalty' accounts for a substantial part of the overall gender earnings gap (Kleven et al. 2019). Goldin (2014) shows that shorter weekly work hours are associated not only with reduced total earnings proportional to the lower labor supply, but also with lower hourly wage rates in a range of highly paid occupations that reward long hours. We hypothesize that in addition to these static effects of lower labor supply on earnings, there may be a dynamic effect where shorter work hours of women slow their on-the-job accumulation of skills and thus reduce their career progression to more highly paid jobs.³⁵

The LinkedIn data unfortunately provide no information on motherhood and do not indicate individuals' work hours in current and past jobs.³⁶ To assess the relationships between skill growth by gender, work hours, and motherhood, we therefore combine information from the LinkedIn profiles with American Community Survey (ACS) data at the level of cells delineated by age, education, and geographic region. Each cell c is defined by a two-year age range, two levels of educational degrees (undergraduate/postgraduate), and nine geographic census divisions. For each such cell, we use the ACS to compute the ratio of average annual

³⁵Cook et al. (2021) observe that female ride-share drivers accumulate fewer hours of work experience per calendar month and therefore exhibit slower growth in productivity than their male counterparts.

³⁶The LinkedIn data only allow to derive the number of months worked since college graduation as a relatively coarse measure of labor supply. Panel A of Appendix Figure A7 indicates that for a given level of potential experience, women have accumulated less actual experience than men on average. Panel B in the figure additionally reports that women are considerably more likely than men to have a low labor-market attachment in the sense that past career breaks lead to a low ratio of actual to potential work experience.

work hours of women versus men in the cell.³⁷ We also use the same source to compute the fraction of women in the cell who live in a household that comprises a child under the age of 18, which provides a proxy for motherhood.³⁸ Finally, we draw on the LinkedIn data to approximate the two-year skill growth for either women or men in the cell by subtracting the average number of skills reported by that group from the average skills of individuals of the same gender, education, and geographic location who are two years older.

Panel A of Figure 2.9 shows the cell-level relationship between skill growth among the females in a cell c , ds_c^f , and skill growth among the men in the cell, ds_c^m . While there is a strong positive correlation between these variables, female skill growth lags male skill growth in most cells, and the slope of the regression line indicates that women on average acquire only 0.82 skills per skill obtained by men. Panel B of Figure 2.9 explores the relationship between this gender gap in skill accumulation and gender differences in labor supply. It fits separate regression lines for worker cells in which the ratio of average female versus average male work hours is relatively high (female/male hours ratio exceeding the median value of 0.82), and cells in which women provide considerably fewer worker hours than men (else). In the cells with relatively high female labor supply, women acquire nearly 0.9 skills per every skill gained by men, whereas in cells with low female labor supply, women obtain only 0.4 skills per added skill of men.

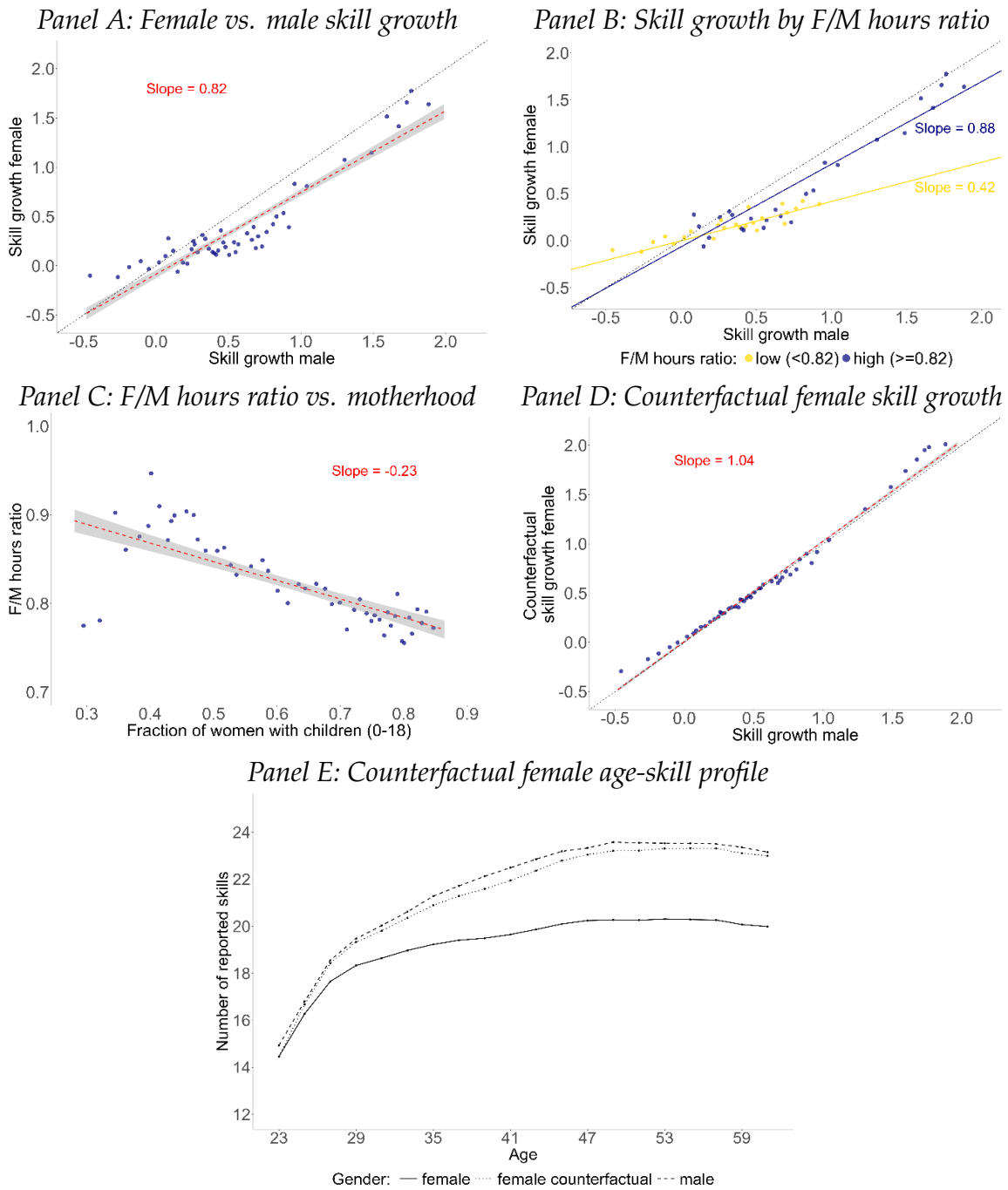
We explore more formally how the relationship between female skill growth ds_c^f and male skill growth ds_c^m depends on the ratio of female vs. male annual work hours in the cell (h_c^f/h_c^m) by estimating the cell-level regression

$$ds_c^f = \alpha + \beta ds_c^m + \gamma \frac{h_c^f}{h_c^m} ds_c^m + \delta \frac{h_c^f}{h_c^m} + \varepsilon_c \quad (2.1)$$

³⁷We compute average female and male work hours for individuals in the cell inclusive of individuals who provide no work hours in a given year, since both reduced hours and spells of non-employment contribute to lower labor-market experience.

³⁸The ACS reports the relationship of every household member to the household head. If a woman is not the household head, it is not always clear whether she is the mother of a child that is present in the household, though motherhood is likely in most cases. We do not simultaneously consider the share of men with children in the household since parenthood is associated with little labor supply change among men (Kleven et al. 2019).

Figure 2.9: Skill accumulation by gender, relative female/male work hours, and motherhood



Notes: Panels A-D are binned scatterplots of 342 cells defined by two-year age bins, two educational degrees, and nine census divisions. Panel A shows the relationship of the growth in the number of LinkedIn skills between women and men. Panel B fits separate regression lines for cells with above and below median female/male (F/M) annual work hours ratio from the American Community Survey (ACS). Panel C shows the relationship between the F/M hours ratio and the fraction of women with children aged 0-18 in ACS data. Panel D indicates the relationship between male skill growth and a counterfactual female skill growth based on an adjusted F/M hours ratio that corrects for the impact of motherhood on work hours suggested by Panel C. Panel E shows actual age-skill profiles for males and females, and a counterfactual profile for females consistent with the counterfactual skill growth of Panel D.

In this regression, a coefficient estimate of $\hat{\gamma} = 1$ would imply that in a cell where women provide 10 percent fewer work hours than men, women's skill growth is 10 percent smaller than that of men. Appendix Table A5 provides estimates. It indicates a coefficient estimate $\hat{\gamma} = 2.28$ (s.e. 0.40) that is positive and significantly larger than unity. This suggests that lower work hours of women relative to men may be associated with a disproportional penalty in terms of skill accumulation. Such a penalty would, for instance, result if firms are more likely to provide training or job tasks that offer learning opportunities to employees who work longer hours, in addition to paying higher hourly wage rates (Goldin 2014).³⁹

We next relate the female-to-male work hours ratio to the fraction of women in the cell who live in a household with an underage child. Panel C of Figure 2.9 shows a highly significant negative association between the proxy for motherhood and the relative labor supply of women. For every 10 percentage-point increase in the share of women with children in the cell, the female-to-male ratio of work hours is 2.3 percentage points lower. If one extrapolates this cell-level relationship to an individual person who becomes mother (i.e., for whom the child variable changes from 0 to 100 percent), then having a child is associated with a 23 percent decline in work hours relative to men. While this relationship indicates only a correlation, its magnitude is consistent with event study evidence on parenthood. Kleven et al. (2024) estimate that parenthood reduces women's labor supply by 25 percent relative to men in North America.

We finally assess how much of the gender skill gap would disappear absent the motherhood-related reduction in female work hours. To construct a counterfactual, motherhood-adjusted skill growth for women, we proceed as follows: First, we compute for every cell a motherhood-adjusted counterfactual female-male hours ratio by adding to the actual hours ratio another 0.23 percentage points for every percentage point of the share of women with children in the cell. Next, we use the estimates for equation 2.1 from Appendix Table A5 to compute the female skill growth for the cell that is consistent with the motherhood-adjusted hours ratio. Panel D of Figure 2.9 shows that absent the motherhood-related reduction in female work hours, female skill accumulation would look strikingly like the skill accumulation of males in the same cell. Finally, we construct a counterfactual age-skill profile for women as the sum of average skills that women reported in the

³⁹There is some evidence for a lower incidence of formal on-the-job training for women relative to men (Barron et al. 1993; KNOKE and ISHIO 1998) and for mothers in particular (Pan et al. 2021).

youngest age cell (age 23-24), plus the cumulative counterfactual cell-level female skill growth for every subsequent higher age. Panel E of Figure 2.9 shows that this counterfactual, which is based on higher female work hours absent motherhood, yields an age-skill profile for women that closely traces men's skill profile at every age. The gender skill gap of Panel A of Figure 2.7, which is small at career start but starts to widen when workers reach their late twenties and thus typical ages of first motherhood for college-educated women in the United States (Nitsche and Brckner 2021), may thus largely be a result of motherhood, which reduces work hours and skill accumulation among women.

2.7.2 Skills and the Gender Earnings Gap

There is a long tradition in labor economics to investigate the contribution of gender differences in human capital to earnings differences between women and men. As women's educational attainment and labor force attachment increased, gender differences in human capital measured by education and experience variables declined to an extent that they can no longer explain a meaningful portion of the remaining gender earnings gap (Goldin 2014; Blau and Kahn 2017; Olivetti et al. 2024).⁴⁰ Given our new finding of a substantial gender skill gap, we next investigate whether our skill measures of human capital account for a larger portion of the earnings differential between women and men.

Since our job- and occupation-based earnings variables provide a metric of the types of jobs that people hold, rather than their actual individual earnings, it is useful to first consider to what extent these job- and occupation-based earnings measures capture a relevant part of the overall gender earnings gap. In the sample of all employed college graduates in the ACS 2019, which we use as comparison set to the LinkedIn data, the raw gender gap in average log annual earnings of full-time, full-year workers is 30.3 log points.⁴¹ The primarily job-based earnings measure of Revelio Labs, which uses information about job title, firm, location, and tenure but *not* gender, shows a 21.7 log point raw gender earnings gap in

⁴⁰In the words of Goldin (2014, p. 1093), "For a long time the gender gap in wages has been viewed as summarizing human capital differences between men's and women's productivity as well as differential treatment of men and women in the labor market. As the grand gender convergence has proceeded, underlying differences between the human capital capabilities of women and men have been vastly reduced and in many cases eliminated."

⁴¹This statistic is based on the wage and salary income of all workers in the ACS 2019 who report having worked at least 51.5 weeks last year and having at least 35 usual weekly work hours.

the LinkedIn sample. This suggests that slightly less than three quarters of the overall gender earnings gap ($21.7/30.3 = 71.7\%$) are due to gender differences in these job-based earnings. Our alternative earnings measure, which assigns to each worker the average ACS-based earnings of college graduates in the worker's occupation, yields a smaller but still sizable raw gender earnings gap of 13.7 log points. These statistics confirm prior research which found that gender differences in occupation (Sloane et al. 2021) and additional job-related characteristics such as industry (Blau and Kahn 2017; Olivetti et al. 2024) or firm wage premia (Goldin et al. 2017) account for a sizable part of the overall gender earnings gap. However, the literature often found it difficult to interpret whether the differential sorting of women and men across jobs reflects job-based discrimination of equally qualified workers, or underlying differences in skills that are not captured by education and experience variables (Blau and Kahn 2017).

Our subsequent regression analysis estimates to what extent different sets of human capital variables can statistically explain the gender differentials in job- and occupation-based earnings, which in turn account for a large portion of the overall gender earnings gap. The upper panel of Table 2.3 regresses workers' log job-based earnings on a female indicator and the different vectors of education, experience, and skill variables that we explored in Table 2.2 above.

The first column in Panel A of Table 2.3, which includes only the female indicator, shows the indicated raw gender gap of 21.7 log points in full-time equivalent job-based earnings for our sample of college graduates. The basic Mincer-style regression in column 2 that controls for five levels of educational degrees and a quartic in potential experience decreases the coefficient estimate of the female indicator only slightly, which confirms the conclusion of prior research that basic human capital proxies explain only a small part ($1 - (-0.196 / -0.217) = 9.7\%$) of the gender earnings gap. The augmented OLS regression model in column 3, which adds a large vector of education and experience variables including detailed indicators for field of study, can account for over a third (35.5%) of the raw earnings gap, and that proportion rises further (43.8%) when using a more flexible random forest specification in column 4.⁴²

⁴²We use the random forest method introduced in **Wager2018**, and extended in Athey et al. (2019), using 750 trees, to estimate the covariate-adjusted difference in earnings between females and males in Table 2.3. This method estimates treatment effects by orthogonalizing the treatment and outcome models, providing point estimates of the treatment parameter while flexibly accounting for nonlinear effects and high-order interactions of the covariates.

Table 2.3: Skills and the gender earnings gap

	Gender	Education/experience			Skills			Combined	
	wage gap (1)	Basic OLS (2)	Detailed OLS (3)	Random forest (4)	Basic OLS (5)	Detailed OLS (6)	Random forest (7)	Detailed OLS (8)	Random forest (9)
A. Job-based earnings (Revelio Labs)									
Female	-0.217 (0.0003)	-0.196 (0.0003)	-0.140 (0.0003)	-0.122 (0.0004)	-0.144 (0.0003)	-0.111 (0.0003)	-0.087 (0.0004)	-0.094 (0.0003)	-0.072 (0.0003)
Reduction vs. raw gap (in percent)		9.7	35.5	43.8	33.6	48.8	59.9	56.7	66.8
B. Occupation-based earnings (ACS)									
Female	-0.137 (0.0002)	-0.129 (0.0002)	-0.082 (0.0002)	-0.072 (0.0003)	-0.098 (0.0002)	-0.061 (0.0002)	-0.046 (0.0003)	-0.048 (0.0002)	-0.034 (0.0002)
Reduction vs. raw gap (in percent)		5.8	40.1	47.4	28.5	55.5	66.4	65.0	75.2
C. Within-occupation variation in job-based earnings (Revelio Labs)									
Female	-0.099 (0.0003)	-0.086 (0.0003)	-0.070 (0.0003)	-0.060 (0.0003)	-0.060 (0.0003)	-0.059 (0.0003)	-0.047 (0.0003)	-0.054 (0.0003)	-0.042 (0.0003)
Reduction vs. raw gap (in percent)		13.1	29.3	39.4	39.4	40.4	52.5	45.5	57.6
<i>Regressors included</i>									
Potential experience (squared)		✓	✓	✓				✓	✓
Actual experience (squared)			✓	✓				✓	✓
Highest degree		✓	✓	✓				✓	✓
Field of study			✓	✓				✓	✓
College ranking			✓	✓				✓	✓
Number of skills					✓	✓	✓	✓	✓
Shares of specific, managerial skills					✓	(✓)	(✓)	(✓)	(✓)
Shares of skills by skill clusters						✓	✓	✓	✓

Notes: Columns 1–3, 5, 6, and 8 are OLS; columns 4, 7, and 9 are random forests. $N = 8,850,314$. Panel A dependent variable: log annual job-based earnings (Revelio Labs). Panel B: log annual earnings at SOC four-digit level (ACS 2018–2019). Panel C: residuals from regressing job-based earnings on SOC four-digit fixed effects. All specifications include an indicator for female; its coefficient is reported as “Female.” “Reduction vs. raw gap” shows the percentage reduction in the female coefficient relative to column 1. Additional regressor definitions as in Table 2. Robust standard errors in parentheses.

These proportions, however, fall well short of the explanatory power of the skill variables. The basic skill model in column 5 which controls for just three skill variables (number of reported skills and shares of specific and managerial skills) already accounts for nearly the same fraction (33.6%) of the gender gap in job-based earnings as the highly detailed regression in column 3 that comprises 42 education and experience variables. Models that consider the distribution of skills across detailed skill clusters explain nearly half (48.8%) of the raw gender gap in the OLS estimation of column 6, and even three-fifths (59.9%) of that gap in the random forest estimation of column 7.

Finally, the combination of the full set of skill, education, and experience variables explains more than half (56.7%) of the raw gender earnings gap in the OLS model of column 8, and fully two thirds (66.4%) of this differential in the more flexible random forest model of column 9.⁴³ These combined models can thus account for only a modestly larger proportion of the raw gender earnings gap than the skills-only specifications in column 6 and 7.

Panel B of Table 2.3 repeats the analysis of the gender earnings gap for the alternative earnings metric that is based on average occupational earnings. The results are qualitatively similar to those obtained with the more detailed earnings measure of Revelio Labs. The regression model with detailed skill variables explains over half (55.5%) of the occupation-based gender earnings differential in the OLS estimation of column 6, and two thirds (66.4%) of that gap in the random forest model of column 7. The combination with detailed education and experience variables in columns 8 and 9 raises these proportions further to nearly two thirds (65.0%) in the OLS specification, and three quarters (75.2%) in the random forest model.

Finally, Panel C of Table 2.3 examines gender differences in earnings within occupations, using as the outcome variable the residuals from a regression of log job-based earnings on four-digit occupation fixed effects.⁴⁴ The simple OLS model in column 5, which includes three basic skill measures—the number of reported skills and the shares of specific and managerial skills—explains nearly forty percent (39.4%) of the raw within-occupation gender earnings gap of 9.9 log points in Revelio Labs' job-based earnings measure. This explanatory power exceeds that of both the basic OLS model with six education and experience variables in column 2 (13.1%) and the much more detailed OLS model with 42 such variables in column 3 (29.3%). In the random forest models, detailed skill variables alone explain more than half of the within-occupation gender gap (52.5 %) in column

⁴³Since gender differences in job-based earnings account for about 72 percent of the overall gender earnings gap among college graduates, the two thirds of the gender gap in job-based earnings explained by the human capital variables in column 9 of Table 2.3 correspond to about half of the overall gender earnings gap ($66.8\% \times 71.7\% = 47.9\%$). This simple quantification may provide only a lower bound for the contribution of human capital variables to the gender earnings gap as it disregards any additional contribution of gender differences in detailed skill, education, and experience variables to earnings differences between women and men within job types.

⁴⁴As shown in Appendix Figure A1, the earnings measure of Revelio Labs has a similar variation across occupations as the ACS-based occupational earnings. Regressions that use average earnings by occupation in the Revelio Labs data as the outcome variable yield a similar pattern of coefficients as the ACS-based measure in Panel B of Table 2.3.

7, and that share rises slightly further (57.6%) when education and experience variables are also included in column 9.

Overall, these results suggest that human capital differences in terms of skills account for a substantial share of the gender earnings gap among college graduates and contribute to gender-based earnings differentials both across and within occupations. Gender differences in human capital thus appear to be a more important contributor to the gender earnings gap than previously suggested by studies that measured human capital with basic education and experience variables. It is however important to note that our findings do not need to imply that gender discrimination in the labor market is less severe than previously thought. Our analysis does not rule out that the observed gender skill gap could arise from discriminatory barriers that limit women's access to training and skill development. The correlational nature of our earnings results also does not preclude the possibility that some men have better access to high-paying jobs than equally qualified women, which may in turn enable them to accumulate more valuable skills through on-the-job experience. However, the fact that recruiters frequently rely on LinkedIn skill listings when searching for candidates suggests that possessing relevant skills is often a prerequisite for accessing such jobs in the first place. What our gender earnings analysis does show is that, while women on average hold considerably lower-paid jobs than men, the extent of their 'underplacement' is substantially reduced when one conditions on their self-reported skills.

2.8 Conclusions

Modern theories of the labor market purport that workers require a multitude of skills to perform a broad range of job tasks, and evidence from job advertisements confirms that employers search for a wide variety of worker skills. Empirical research, however, mostly reverts to proxying for individual human capital by educational attainment and work experience variables that capture inputs into human capital production rather than the skills that result from that production.

Rich data from online professional profiles on the LinkedIn platform provide us with a new opportunity to study the role of human capital in the labor market. Users of the platform self-report CV information including skills to signal their qualifications to potential employers. We gather this information for 8.85 million college graduates in the U.S. labor market for whom we also observe detailed

information on education and experience, and can infer job-based earnings from information such as job titles and occupations. We aggregate the thousands of reported skill strings into 48 skill clusters using a word association algorithm and use the ensuing database to perform three sets of analyses.

First, we show that self-reported skills correlate sensibly with experience and education variables that capture inputs into skill production. Older and more experienced workers report a larger number of skills. Their skill sets also contain a larger share of skills that are occupation-specific rather than widespread across many occupations, as well as a larger share of managerial skills. These patterns are consistent with the notion that young workers primarily obtain general skills through education while older workers additionally acquire specific and managerial skills through work experience and post-university training. Skill patterns also vary across workers with different types of education, where studies in STEM, health, and law fields and more advanced degrees provide a higher fraction of occupation-specific skills.

Second, there is a clear tendency for workers with more reported skills to be employed in higher-paying jobs. Many specific and managerial skills are particularly strongly associated with higher job-based earnings. While one may be concerned that self-reported skills provide only a noisy measure of human capital, we show that skills explain a much higher fraction of earnings variation than basic education and experience variables, and even have higher explanatory power than an unusually detailed vector of education and experience variables that accounts for field of study, college quality, and actual post-college work experience. Indeed, a sizable part of the well-known relationship between education, experience, and earnings appears to be mediated by the skills observed on LinkedIn profiles.

Third, we document a substantial gender gap in self-reported skills between female and male college graduates, particularly at older ages. At labor market entry, women report slightly fewer skills than men and are less likely to list specific or managerial skills. However, much of this initial gap in skill quantity and composition vanishes when comparing men and women within the same occupations. Among workers in their late twenties to early forties, a notable divergence emerges: the number of reported skills grows significantly more for men than for women, even within the same occupational categories. We show that this pattern aligns with the effects of reduced labor supply following motherhood, which limits women's opportunities for on-the-job skill accumulation. Importantly,

these gender differences in skill profiles explain a substantial portion of the gender earnings gap—far more than traditional variables such as educational degrees or years of experience, which now show minimal gender disparities.

Overall, our analysis underscores that conventional measures of human capital based on education and work experience may not sufficiently reflect the breadth and complexity of workers' skills. Our findings suggest that LinkedIn skill data offer a valuable complement to traditional metrics, enabling a richer understanding of human capital differences across diverse groups in the labor market.

2.A Appendix A: Additional Tables and Figures

Table A1: Most frequent skill strings in each skill cluster

Skill cluster	Most frequent skill strings
Specific skills	
Accounting, financial reporting	Accounting, financial reporting, auditing, financial accounting, accounts payable
AutoCAD, SolidWorks	AutoCad, SolidWorks, CAD, mechanical engineering, finite element analysis
Biotechnology, molecular biology	Biotechnology, molecular biology, chemistry, cell culture, PCR
Clinical research, medical devices	Clinical research, medical devices, pharmaceutical industry, clinical trials, oncology
Engineering, project engineering	Engineering, project engineering, petroleum, gas, electrical engineering
Healthcare, hospitals	Healthcare, hospitals, healthcare management, nursing, EMR
Insurance, banking	Insurance, banking, credit, loans, property
Java enterprise edition, Jira	Java Enterprise Edition, Jira, JSP, test automation, Hibernate
Java, Matlab	Java, Matlab, C++, Python, C
Legal research/writing	Legal research, legal writing, litigation, civil litigation, courts
Mobile devices/applications	Mobile devices, mobile applications, Objective-C, iOS, iOS development
Real estate, investment properties	Real estate, investment properties, real estate transactions, residential homes, commercial real estate
Revit, SketchUp	Revit, SketchUp, product design, architecture, interior design
SQL, software development	SQL, software development, Linux, agile methodologies, Microsoft SQL Server
Telecom., network security	Telecommunications, network security, wireless, Cisco technologies, VoIP
Windows server, disaster recovery	Windows server, disaster recovery, servers, data center, VMware
General managerial skills	
Analysis, financial analysis	Analysis, financial analysis, finance, risk management, financial modeling
Business analysis/process improv.	Business analysis, business process improvement, integration, change management, SDLC
Coaching, leadership development	Coaching, leadership development, curriculum development, staff development, curriculum design

Notes: The table shows the five most commonly reported raw skills for the 48 skill clusters.

Table A1: Most frequent skill strings in each skill cluster (continued)

Skill cluster	Most frequent skill strings
Marketing, social media marketing	Marketing, social media marketing, marketing strategy, advertising, marketing communications
Program mgmt., security clearance	Program management, security clearance, military, DoD, proposal writing
Project management, budgets	Project management, budgets, project planning, policy, government
Sales, strategic planning	Sales, strategic planning, team leadership, account management, strategy
General non-managerial skills	
Access, software documentation	Access, software documentation, Visio, SAP, ERP
Communication, problem solving	Communication, problem solving, organization, organization skills, organizational leadership
Customer satisfaction/retention	Customer satisfaction, customer retention, automotive, call centers, customer experience
Data analysis, databases	Data analysis, databases, statistics, SPSS, qualitative research
Editing, public relations	Editing, public relations, writing, blogging, creative writing
Energy, sustainability	Energy, sustainability, environmental awareness, environmental science, renewable energy
English, Spanish	English, Spanish, education, history, French
Excel, customer relations	Excel, customer relations, Word, planning, IIS
Food, hospitality	Food, hospitality, beverage, restaurants, hospitality management
HTML, JavaScript	HTML, JavaScript, CSS, web design, WordPress
Info. technology, lean six sigma	Information technology, lean six sigma, people management, people development, IT solutions
Inventory/operations management	Inventory management, operations management, logistics, supply chain management, purchasing
Microsoft Office, customer service	Microsoft Office, customer service, leadership, Microsoft Excel, management
Photoshop, Adobe CS	Photoshop, Adobe Creative Suite, graphic design, photography, InDesign
Process improv., cross-func. team lead.	Process improvement, cross-functional team leadership, manufacturing, continuous improvement, Six Sigma
Public speaking, research	Public speaking, research, teaching, community outreach, nonprofits
Recruiting, human resources	Recruiting, human resources, employee relations, interviews, employee benefits

Notes: The table shows the five most commonly reported raw skills for the 48 skill clusters.

Table A1: Most frequent skill strings in each skill cluster (continued)

Skill cluster	Most frequent skill strings
Retail, forecasting	Retail, forecasting, merchandising, visual merchandising, pricing
Security, emergency management	Security, emergency management, first aid, criminal justice, homeland security
Social media/networking	Social media, social networking, Facebook, sports, wellness
Testing, quality assurance	Testing, quality assurance, SharePoint, technical writing, FDA
U.s, software dev. life cycle	U.s, software development life cycle, relationship building, waterfall methodologies, communications
Video production/editing	Video production, video editing, music, Final Cut Pro, video
Windows, troubleshooting	Windows, troubleshooting, networking, technical support, system administration

Table A2: Sample selection

Filtering step	Observations	Share			Age	Female	College degree
All LinkedIn profiles, U.S. workers, 2019	61,850,913	1.00	1.00	1.00	37.75	0.48	0.44
Non-missing baseline data							
Non-missing gender	57,835,032	0.94	0.94	0.94	37.86	0.48	0.43
Non-missing occupation	57,834,793	0.94	1.00	1.00	37.86	0.48	0.43
Non-missing state	55,418,826	0.90	0.96	0.96	37.88	0.48	0.44
Non-missing education data							
Any education stage	43,450,459	0.70	0.78	0.78	37.88	0.48	0.56
Mapped highest degree	33,673,272	0.54	0.78	0.60	37.88	0.49	0.73
Valid age and experience indicators							
Highest degree with valid date	28,343,736	0.46	0.84	0.51	37.88	0.48	0.86
Graduation date before scrape date	25,922,593	0.42	0.92	0.46	37.88	0.47	0.94
Imputed age at first job at least 14	24,320,432	0.39	0.94	0.43	38.39	0.48	0.94
First job at most 5 years after graduation	20,658,665	0.33	0.85	0.40	35.92	0.48	0.95
Age range and college degree							
Age 23–64	18,755,466	0.30	0.91	0.41	36.08	0.48	0.97
At least Associate’s degree	18,136,202	0.29	0.97	0.44	36.17	0.48	1.00
Non-missing skill data							
At least one reported skill	8,950,793	0.14	0.49	0.38	37.46	0.46	1.00
No non-English skill clusters	8,850,314	0.14	0.99	1.00	37.47	0.46	1.00

Notes: Profile characteristics of the U.S. workforce on LinkedIn in 2019 after different sample selection steps. Column (2) reports the share retained relative to the baseline sample, column (3) the share retained relative to the previous step, and column (4) the share retained if the given step were applied to the baseline sample. Columns (5)–(7) report mean age, fraction female, and fraction with a college degree at each step.

Table A3: Skill count and composition by experience and education: Regressions

Dependent variable:	Number of reported skills (1)	Share of skill domain in total skills		
		General non-managerial (2)	General managerial (3)	Specific (4)
Experience				
Potential experience	-2.135 (0.037)	-0.176 (0.0009)	0.088 (0.0007)	0.088 (0.0008)
Potential experience ²	1.762 (0.081)	-0.390 (0.002)	-0.175 (0.002)	-0.214 (0.002)
Actual experience	7.755 (0.040)	-0.107 (0.001)	0.078 (0.0007)	0.029 (0.0009)
Actual experience ²	-12.146 (0.097)	0.118 (0.002)	-0.095 (0.002)	-0.023 (0.002)
Highest degree (Ref: associate)				
Bachelor	1.794 (0.018)	-0.050 (0.0004)	0.040 (0.0003)	0.010 (0.0004)
Master	2.847 (0.018)	-0.055 (0.0004)	0.045 (0.0003)	0.010 (0.0004)
Professional degree	-0.653 (0.027)	-0.187 (0.0007)	-0.044 (0.0005)	0.231 (0.0008)
Doctorate	1.992 (0.027)	-0.070 (0.0006)	-0.054 (0.0004)	0.124 (0.0006)
Field of study (Ref: arts/humanities)				
Business	0.150 (0.015)	-0.226 (0.0003)	0.133 (0.0003)	0.093 (0.0003)
Health/medicine	-3.099 (0.020)	-0.223 (0.0005)	-0.073 (0.0003)	0.296 (0.0005)
Law	-0.218 (0.032)	-0.234 (0.0008)	0.011 (0.0005)	0.223 (0.0009)
Missing	0.486 (0.021)	-0.173 (0.0005)	0.039 (0.0004)	0.134 (0.0004)
Other majors	-0.485 (0.024)	-0.069 (0.0005)	0.037 (0.0004)	0.031 (0.0004)
Social sciences	-0.750 (0.016)	-0.059 (0.0004)	0.060 (0.0003)	-0.0004 (0.0002)
STEM	0.349 (0.016)	-0.240 (0.0003)	-0.009 (0.0003)	0.249 (0.0003)
College ranking (Ref: Top 10)				
11–50	0.679 (0.036)	0.041 (0.0009)	-0.041 (0.0008)	0.0001 (0.0008)
51–100	0.482 (0.035)	0.048 (0.0008)	-0.051 (0.0007)	0.002 (0.0008)
101–200	0.857 (0.035)	0.069 (0.0008)	-0.060 (0.0007)	-0.008 (0.0008)
201–500	0.686 (0.033)	0.082 (0.0008)	-0.068 (0.0007)	-0.015 (0.0007)
501–1000	0.413 (0.033)	0.106 (0.0008)	-0.083 (0.0007)	-0.022 (0.0007)
No US rank	1.299 (0.034)	0.088 (0.0008)	-0.070 (0.0007)	-0.019 (0.0008)
Non-US college	1.888 (0.039)	0.007 (0.0009)	-0.049 (0.0008)	0.043 (0.0009)
Missing college	1.238 (0.035)	0.094 (0.0008)	-0.081 (0.0007)	-0.013 (0.0008)
Observations	8,850,314	8,850,314	8,850,314	8,850,314
R ²	0.052	0.228	0.180	0.221

Notes: OLS regressions. Dependent variable indicated in column header. Regressions include fixed effects for scrape months. Robust standard errors in parentheses.

Table A4: Alternative metrics of human capital and job-based earnings

	Education/experience		Skills		Combined
	Basic (1)	Detailed (2)	Basic (3)	Detailed (4)	Detailed (5)
Experience					
Potential experience	2.939 (0.006)	1.164 (0.014)			0.288 (0.013)
Potential experience ² /100	-4.743 (0.015)	-3.590 (0.030)			-1.370 (0.029)
Actual experience		2.110 (0.015)			1.129 (0.014)
Actual experience ² /100		-1.665 (0.035)			-0.477 (0.034)
Highest degree					
Bachelor	26.236 (0.058)	20.933 (0.064)			16.576 (0.060)
Master	32.330 (0.061)	29.099 (0.067)			24.318 (0.064)
Professional degree	56.310 (0.086)	54.889 (0.105)			49.973 (0.110)
Doctorate	41.277 (0.089)	40.419 (0.098)			40.097 (0.099)
Field of study					
Architecture and related services		17.643 (0.173)			8.793 (0.178)
Biological and biomedical sciences		5.157 (0.176)			3.164 (0.171)
Business, management, marketing		22.549 (0.161)			10.157 (0.157)
Communication, journalism, related programs		10.566 (0.169)			3.601 (0.166)
Computer+information sciences, support serv.		33.734 (0.167)			11.085 (0.167)
Education		-12.305 (0.171)			-5.425 (0.169)
Engineering		27.933 (0.164)			14.530 (0.164)
Engineering/engin.-rel. tech./technicians		12.737 (0.198)			3.219 (0.191)
Family and consumer sciences/human sciences		-8.005 (0.290)			-5.165 (0.280)
Health professions and related programs		9.470 (0.168)			4.896 (0.167)
History		3.985 (0.216)			2.053 (0.207)
Homeland security, law enforcement, firefighting		-1.473 (0.205)			-2.208 (0.201)
Legal professions and studies		13.395 (0.195)			4.703 (0.195)
Liberal arts+sciences, general studies, humanities		2.064 (0.186)			-0.091 (0.180)
Mathematics, statistics, and physical sciences		15.819 (0.185)			7.230 (0.178)
Missing field of study		17.359 (0.169)			8.158 (0.164)
Multi/interdisciplinary studies		10.539 (0.207)			6.110 (0.199)
Natural resources and conservation		5.199 (0.251)			1.683 (0.245)
Other field of study		7.530 (0.191)			4.747 (0.184)
Parks, recreation, leisure, fitness, kinesiology		-3.461			-2.906

(continued on next page)

2. Multidimensional Skills on LinkedIn Profiles: Measuring Human Capital and the Gender Skill Gap

(continued)	Education/experience		Skills		Combined
	Basic (1)	Detailed (2)	Basic (3)	Detailed (4)	Detailed (5)
Philosophy and religious studies		(0.238) 7.112			(0.233) 3.409
Psychology		(0.259) -2.165			(0.247) -1.500
Public administration, social service professions		(0.180) -4.728			(0.175) -2.576
Social sciences		(0.197) 14.347			(0.192) 7.128
Theology and religious vocations		(0.180) -3.699			(0.174) -0.452
Visual and performing arts		(0.269) 5.770			(0.261) 2.672
		(0.179)			(0.177)
College ranking					
11–50		-11.539			-8.395
		(0.135)			(0.127)
51–100		-18.455			-13.495
		(0.132)			(0.124)
101–200		-21.144			-15.541
		(0.132)			(0.123)
201–500		-27.229			-20.059
		(0.127)			(0.119)
501–1000		-33.450			-24.643
		(0.129)			(0.121)
No US rank		-31.329			-23.609
		(0.129)			(0.121)
Non-US college		-17.793			-16.524
		(0.144)			(0.135)
Missing college		-33.937			-25.798
		(0.130)			(0.122)
Number of skills					
Number of skills			0.588	0.408	0.319
			(0.001)	(0.001)	(0.001)
Shares of specific and managerial skills					
Specific skills				0.536	
				(0.0005)	
Managerial skills				0.614	
				(0.0007)	
Shares of skills by skill clusters					
<i>Specific skills</i>					
Mobile devices/applications				1.635	1.349
				(0.012)	(0.011)
SQL, software development				1.251	0.960
				(0.006)	(0.006)
Windows server, disaster recovery				1.276	0.948
				(0.007)	(0.006)
Telecom., network security				1.088	0.800
				(0.007)	(0.006)
Java enterprise edition, Jira				0.945	0.717
				(0.007)	(0.006)
Java, Matlab				1.183	0.709
				(0.006)	(0.006)
Clinical research, medical devices				1.168	0.706
				(0.006)	(0.006)
Accounting, financial reporting				0.988	0.695
				(0.006)	(0.006)
Revit, SketchUp				0.984	0.690
				(0.006)	(0.006)
Healthcare, hospitals				0.919	0.663
				(0.006)	(0.006)

(continued on next page)

(continued)	Education/experience		Skills		Combined
	Basic (1)	Detailed (2)	Basic (3)	Detailed (4)	Detailed (5)
Legal research/writing				1.273 (0.006)	0.655 (0.006)
Engineering, project engineering				0.906 (0.006)	0.648 (0.006)
AutoCAD, SolidWorks				0.923 (0.006)	0.632 (0.006)
Real estate, investment properties				0.852 (0.006)	0.581 (0.006)
Insurance, banking				0.811 (0.006)	0.570 (0.006)
Biotechnology, molecular biology				0.802 (0.006)	0.424 (0.006)
<i>General managerial skills</i>					
Analysis, financial analysis				1.572 (0.006)	1.154 (0.006)
Business analysis/process improv.				1.488 (0.006)	1.033 (0.006)
Sales, strategic planning				1.298 (0.006)	0.966 (0.006)
Marketing, social media marketing				1.215 (0.006)	0.931 (0.006)
Project management, budgets				1.104 (0.006)	0.769 (0.006)
Program mgmt., security clearance				0.870 (0.006)	0.563 (0.006)
Coaching, leadership development				0.559 (0.006)	0.327 (0.006)
<i>General non-managerial skills</i>					
Info. technology, lean six sigma				1.555 (0.016)	1.230 (0.014)
HTML, JavaScript				1.003 (0.006)	0.860 (0.006)
U.s, software dev. life cycle				1.016 (0.016)	0.851 (0.014)
Process improv., cross-func. team lead.				1.124 (0.006)	0.730 (0.006)
Testing, quality assurance				0.837 (0.007)	0.640 (0.006)
Recruiting, human resources				0.858 (0.006)	0.584 (0.006)
Energy, sustainability				0.915 (0.007)	0.562 (0.006)
Access, software documentation				0.752 (0.007)	0.521 (0.007)
Data analysis, databases				0.872 (0.006)	0.508 (0.006)
Video production/editing				0.671 (0.006)	0.500 (0.006)
Inventory/operations management				0.695 (0.007)	0.486 (0.006)
Editing, public relations				0.711 (0.006)	0.461 (0.006)
Social media/networking				0.559 (0.006)	0.412 (0.006)
Other Skills				0.658 (0.006)	0.412 (0.006)
Photoshop, Adobe CS				0.551 (0.006)	0.408 (0.006)
Public speaking, research				0.685 (0.006)	0.407 (0.005)

(continued on next page)

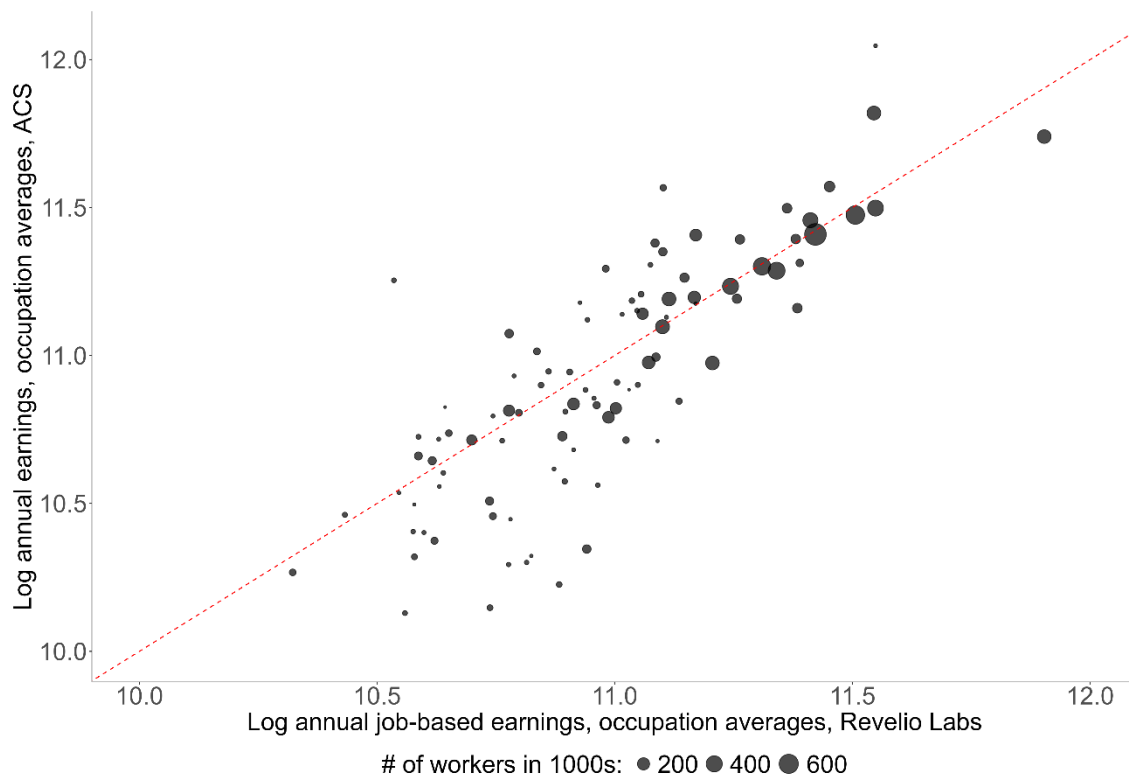
(continued)	Education/experience		Skills		Combined
	Basic (1)	Detailed (2)	Basic (3)	Detailed (4)	Detailed (5)
Retail, forecasting				0.589 (0.007)	0.397 (0.006)
Food, hospitality				0.532 (0.007)	0.385 (0.006)
Windows, troubleshooting				0.396 (0.007)	0.384 (0.006)
Security, emergency management				0.492 (0.007)	0.369 (0.006)
English, Spanish				0.582 (0.007)	0.357 (0.006)
Excel, customer relations				0.646 (0.010)	0.335 (0.009)
Microsoft Office, customer service				0.333 (0.006)	0.268 (0.006)
Communication, problem solving				0.043 (0.007)	0.071 (0.006)
Fit					
Adj. R ²	0.142	0.235	0.172	0.253	0.325

Notes: $N = 8,850,314$ LinkedIn profiles. The dependent variable is the log of job-based annual earnings imputed by Revelio Labs' proprietary salary model. All regressions are OLS. The model in column (4) provides the coefficient estimates shown in Figure 6. Columns (1)–(5) use the same regressor sets as columns (1), (2), (4), (5), and (7) of Table 2, but are estimated on the full sample rather than a 70% subsample. Potential experience is the time between college graduation and scrape month (years); actual experience is the cumulative time in job spells since graduation (years). Number of skills is the count of skill strings on a profile. "Shares of specific and managerial skills" and "shares by skill clusters" refer to the fraction of skill strings in the respective domain/cluster. Omitted categories: *Associate* (highest degree), agricultural/animal/plant science (field of study), *Top 10* (college ranking), general skills (skill domain shares), and customer satisfaction/retention (skill cluster shares). All regressions include scrape-month fixed effects. Robust standard errors in parentheses.

Table A5: Female skill growth as a function of male skill growth and the female/male work hours ratio

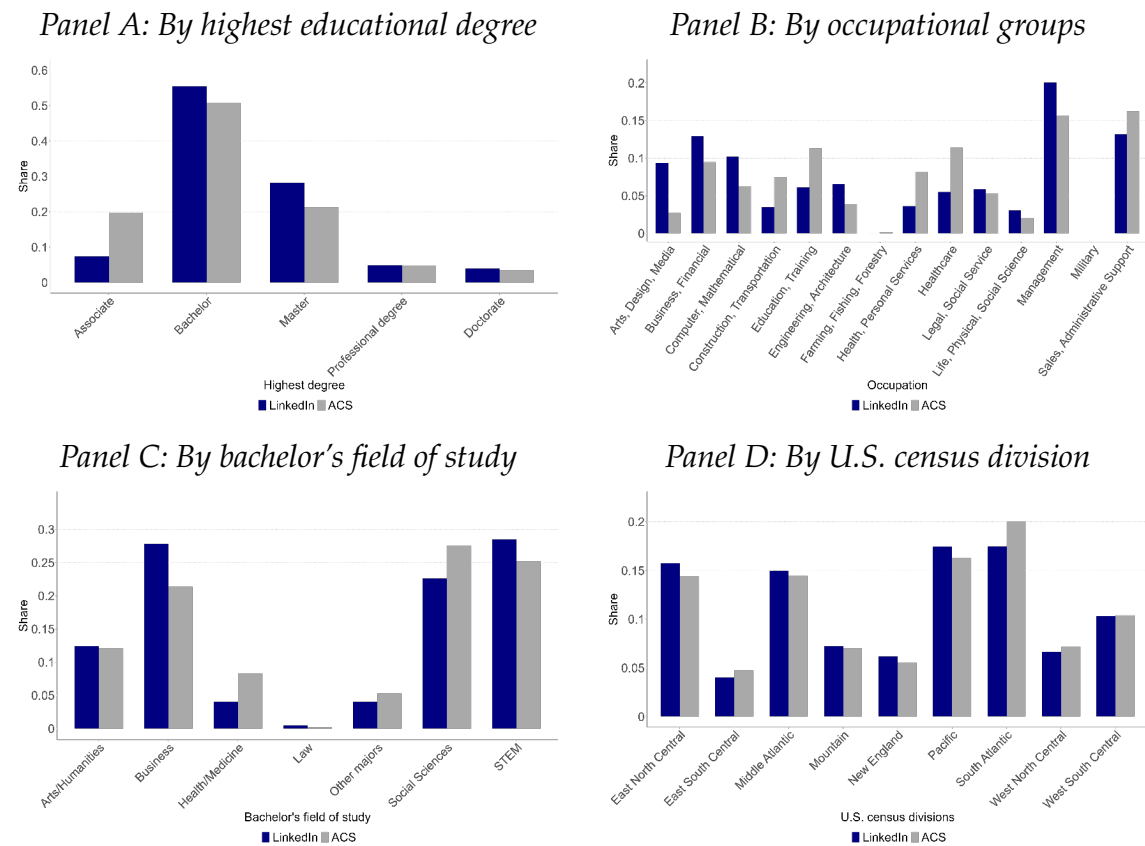
	(1)	(2)
Male skill growth	0.820 (0.036)	-1.303 (0.351)
Female/male work hours ratio × Male skill growth		2.278 (0.403)
Female/male work hours ratio		0.276 (0.360)
Observations	342	342
R^2	0.750	0.821

Notes: Regression of female skill growth on male skill growth, the female/male annual work hours ratio, and their interaction. Cell-level regressions with cells defined by two-year age bins, two educational degrees, and nine census divisions. Female and male skill growth is approximated by subtracting the average number of LinkedIn skills reported by a gender-age-education-geography group from the average skills of individuals of the same gender, education, and geographic location who are two years older. The ratio of average annual work hours of women versus men is computed from pooled 2018–2019 ACS data and based on both individuals with zero and positive work hours. Robust standard errors in parentheses.

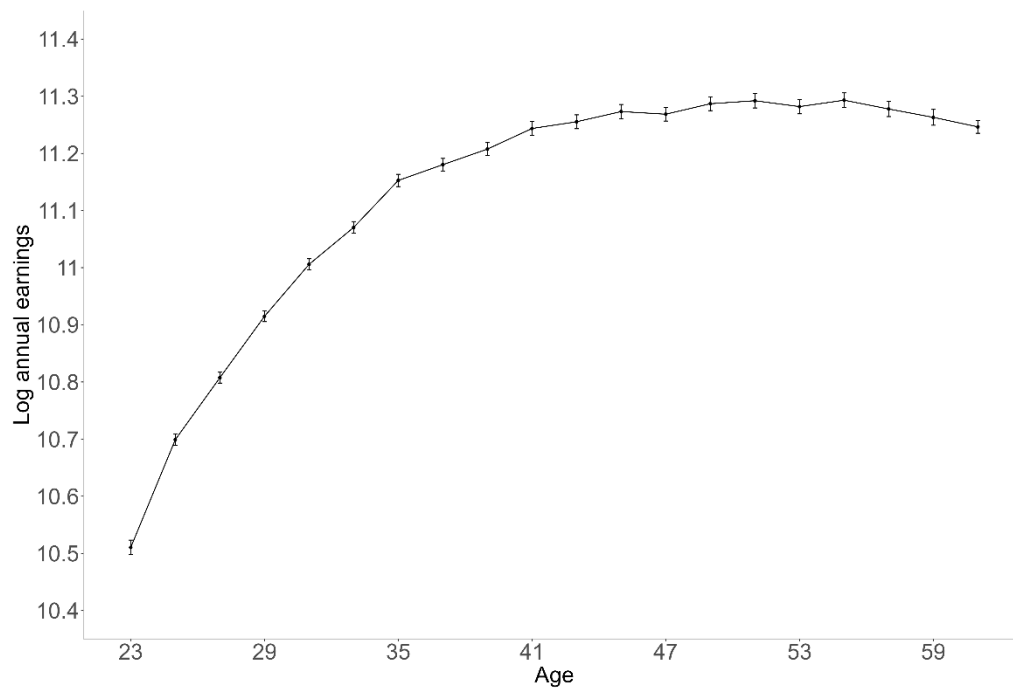
Figure A1: Job-based (Revelio Labs) and occupation-based (ACS) earnings measures

Notes: The x-axis indicates the average log annual job-based earnings that Revelio Labs computes for all employees of the 2019 LinkedIn sample, aggregated to 95 occupations. The y-axis indicates the average log annual earnings of the corresponding occupation in the 2018–2019 American Community Survey (ACS).

Figure A2: Comparison of LinkedIn sample to ACS



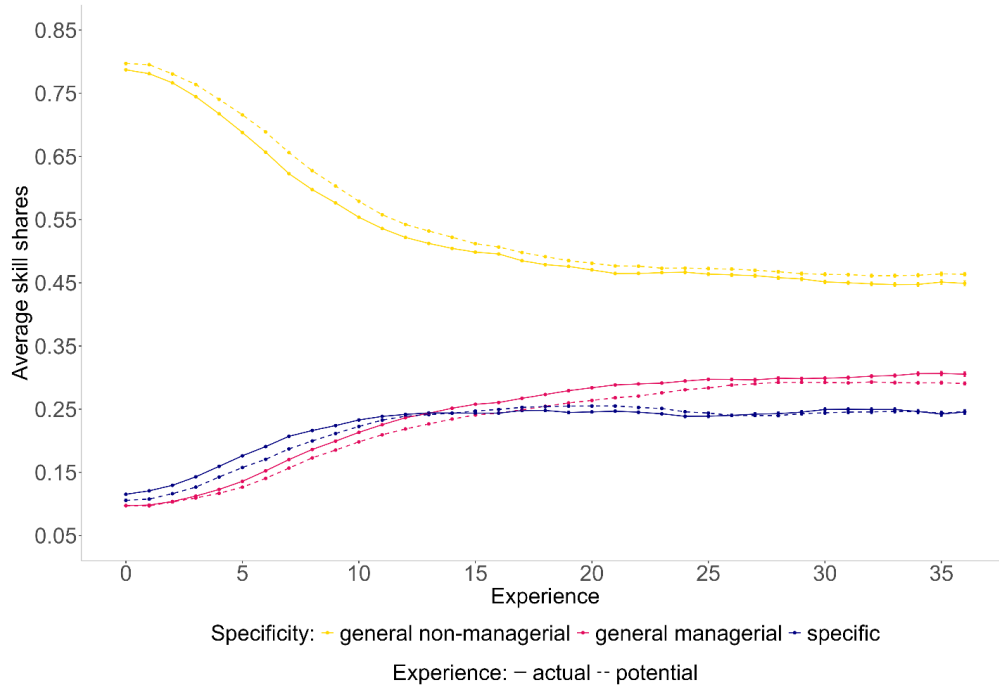
Notes: The figures compare the LinkedIn estimation sample (blue) to the employed college graduates aged 23-64 in the American Community Survey (ACS) 2019 (grey). The ACS 1-percent sample contains 6.5m employed college graduates. Panel B classifies occupation groups as aggregates of SOC two-digit codes as in Deming and Noray (2020). The occupation groups 'military' and 'farming, fishing, forestry' represent only trivial fractions of the estimation sample and total US employment. The statistics in Panel C are based only on individuals with at least a Bachelor's degree and omit the 7 percent of individuals in our sample whose highest qualification is an Associate's degree. Field of study is classified at the CIP two-digit level and aggregated into seven categories.

Figure A3: Age-earnings profile for college graduates

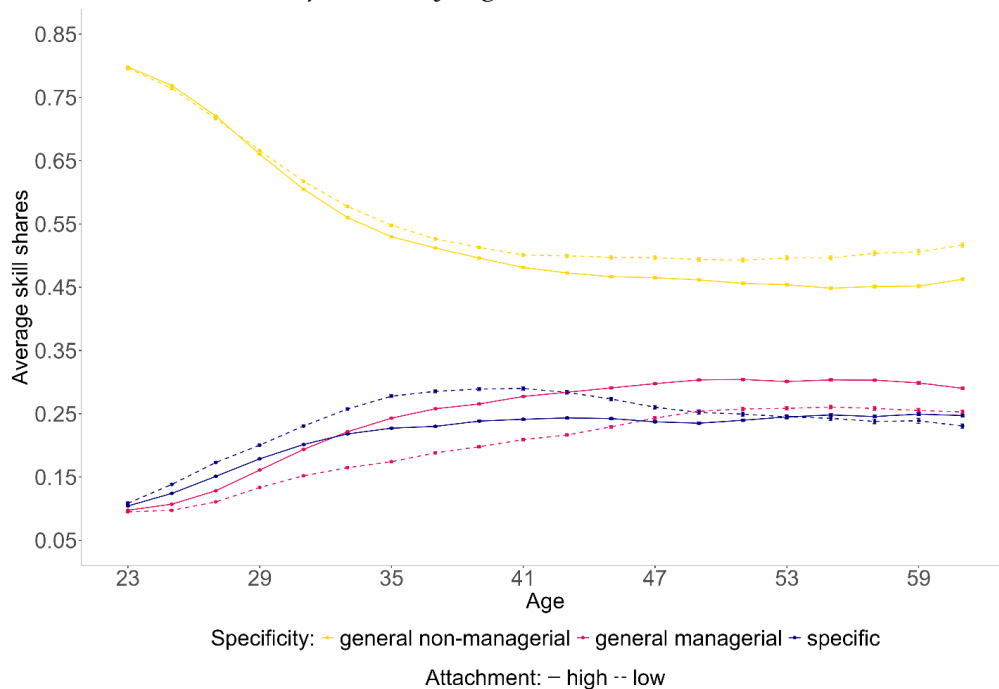
Notes: Average log annual earnings of employed college graduates aged 23-64 in the American Community Survey (ACS) 2019 by age (two-year age bins). Error bars depict 99 percent confidence intervals.

Figure A4: Skill composition by experience and labor-market attachment

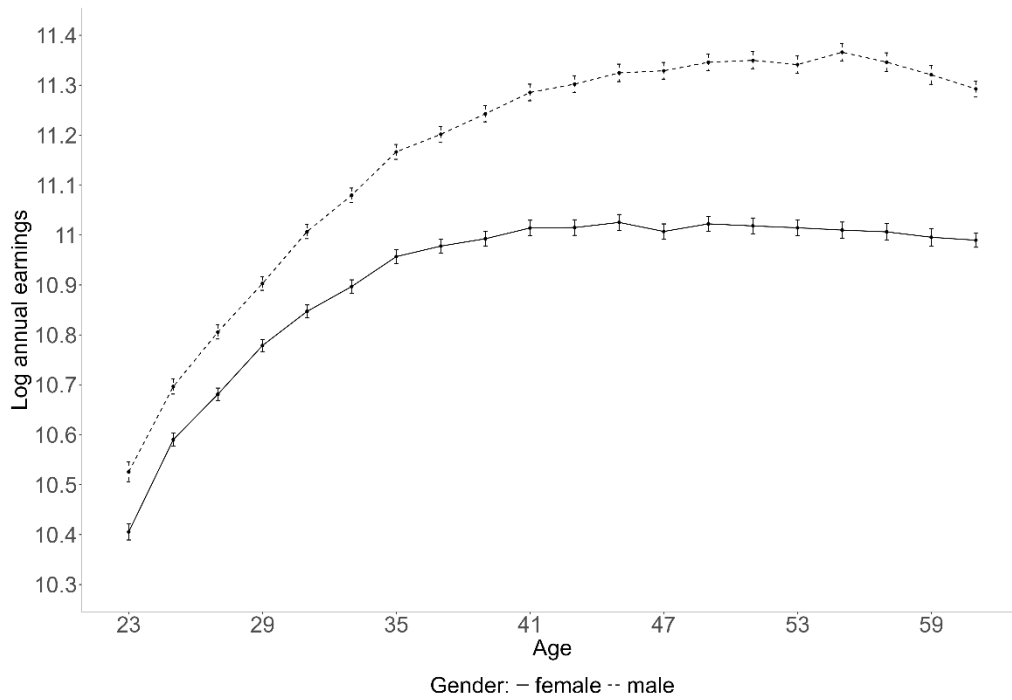
Panel A: Skill composition by actual and potential work experience



Panel B: Skill composition by high and low labor-market attachment



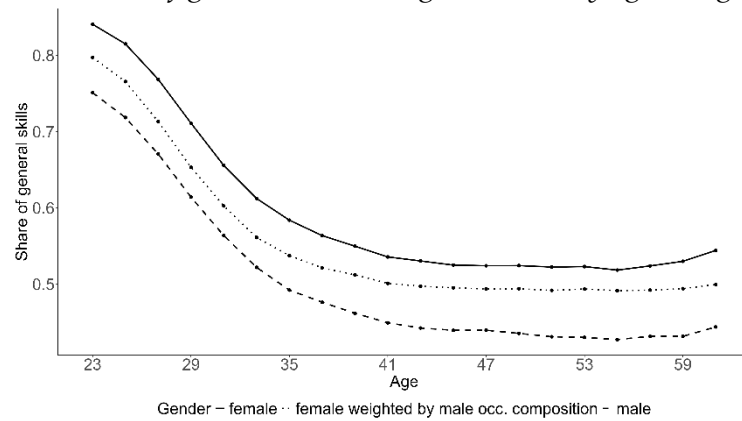
Notes: Panel A reports the average fractions of general non-managerial, general managerial, and occupation-specific skills by potential and actual work experience. Panel B reports the average fractions of general non-managerial, general managerial, and occupation-specific skills by for workers with low labor force attachment (defined as a ratio of actual to potential experience below 0.8 and cumulative non-employment spells of at least half a year) and for workers with high attachment (everyone else). Error bars depict 99 percent confidence intervals.

Figure A5: Age-earnings profiles for college graduates by gender

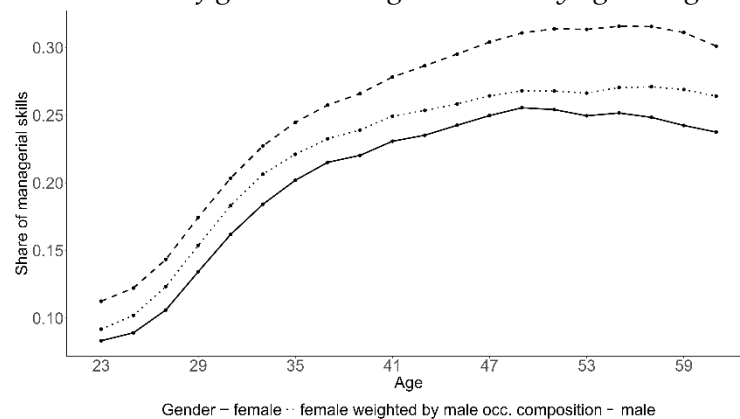
Notes: Average log annual earnings of employed college graduates aged 23-64 in the American Community Survey (ACS) 2019 by age (two-year age bins) and gender. Error bars depict 99 percent confidence intervals.

Figure A6: Skill composition by age and gender conditional on occupation

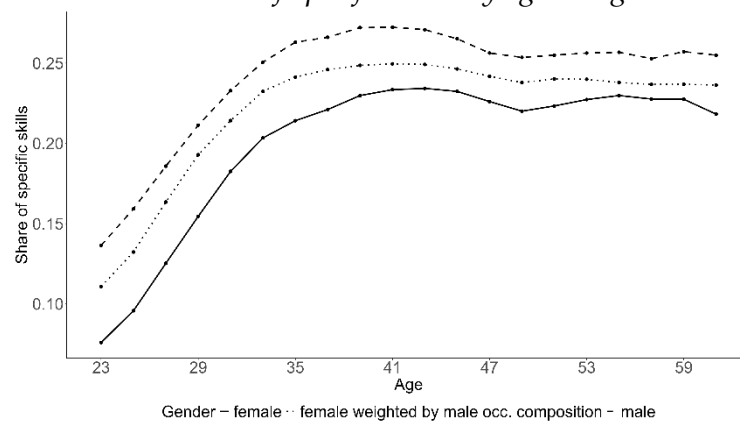
Panel A: Share of general non-managerial skills by age and gender



Panel B: Share of general managerial skills by age and gender

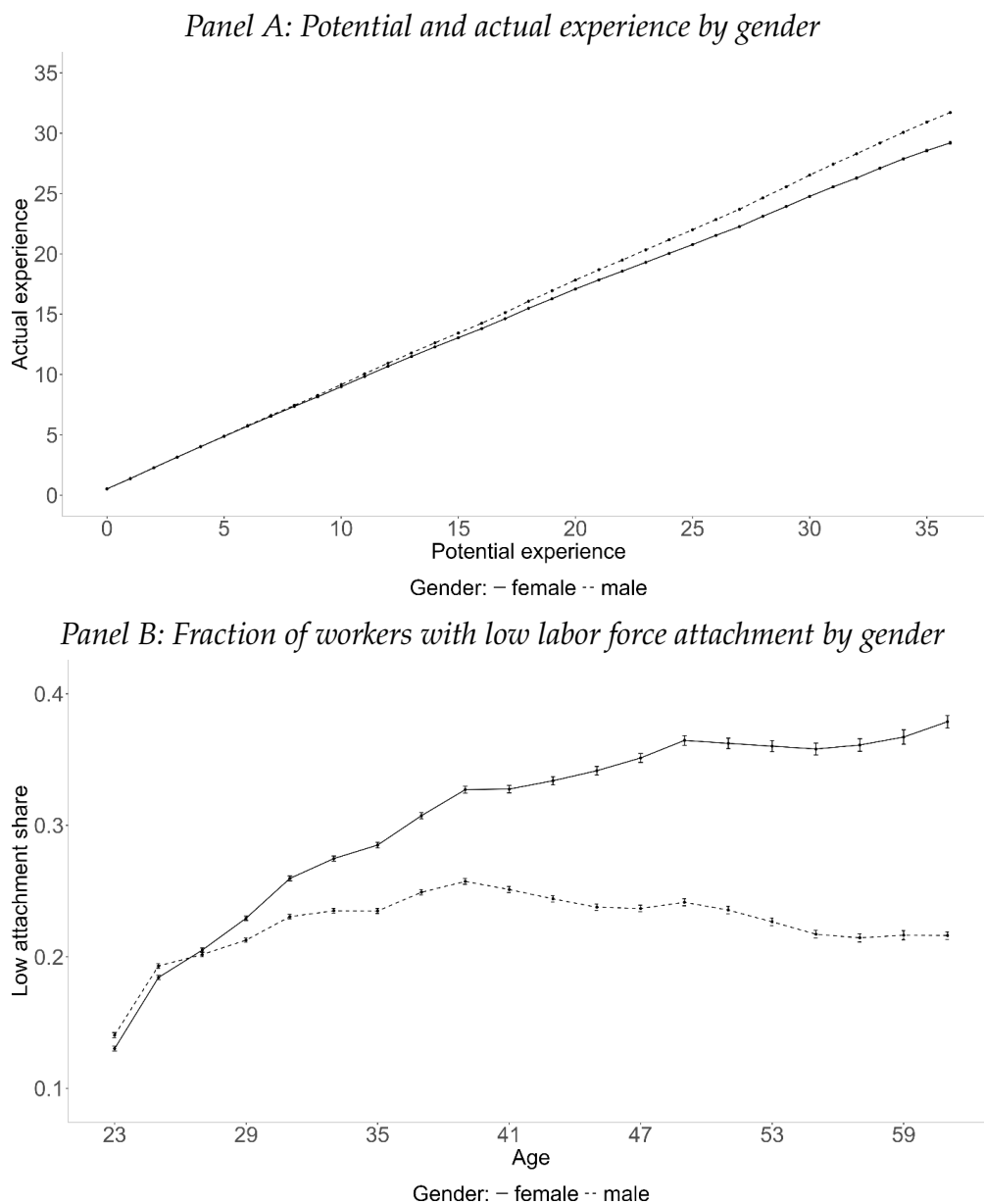


Panel C: Share of specific skills by age and gender



Notes: The figure shows the average fractions of general non-managerial, general managerial, and occupation-specific skills, respectively, by age and gender. The data series ‘female weighted by male occ. composition’ is based on the average number of skills (in Panel A) or average shares of general non-managerial, general managerial, and specific skills for women of a given age group and detailed occupation, weighted by the employment share of each occupation among the men of the same age group. The difference between this data series and the male series indicates within-occupation differences in skills across gender, while the difference with regard to the female series indicates between-occupation differences in skills across gender. Two-year age bins; last bin combines ages 61-64.

Figure A7: Potential experience, actual experience, and labor force attachment by gender



Notes: Low attachment share refers to the share of workers with low labor force attachment, defined as a ratio of actual to potential experience below 0.8 and cumulative non-employment spells of at least half a year. Error bars depict 99 percent confidence intervals.

2.B Appendix B: Results for LinkedIn Users with Recently Updated Profiles

The LinkedIn data may be inaccurate for users who do not regularly update their profiles. While we do not observe the date when a user last modified any profile information, we are able to observe whether a user recently changed to a job at a new firm, in which case the user may also have updated other relevant profile information such as skills prior to applying for the new job.

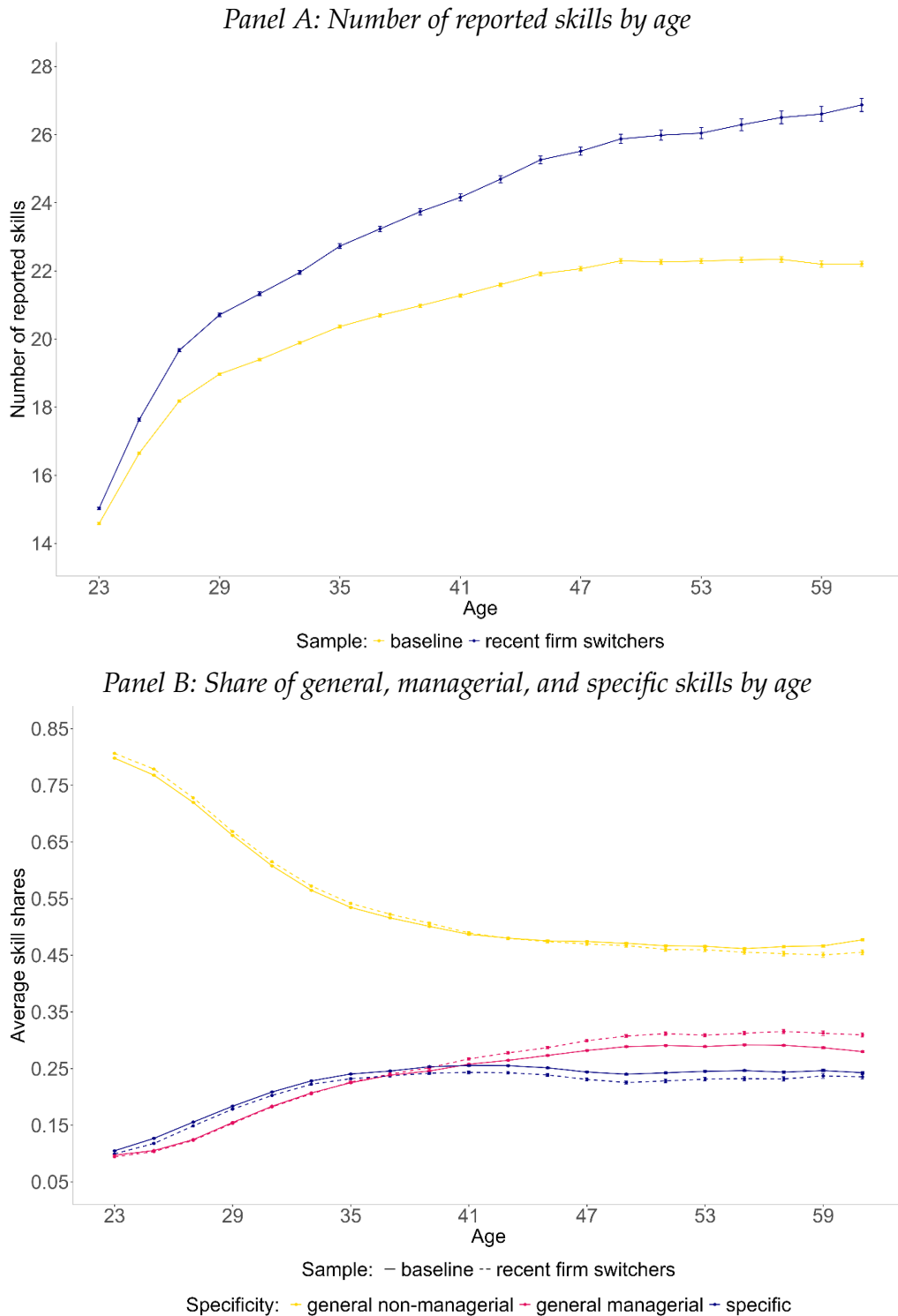
In this appendix, we study the robustness of key findings of our analysis to constraining the LinkedIn sample to users who changed firms within the last two years (in 2017 to 2019).⁴⁵ The resulting subset of profiles will likely be more up to date, but also less representative of the college-educated workforce at large. This is the case because a sample of recent firm switchers will oversample individuals who have an elevated job-to-job mobility, such as early-career workers. Nevertheless, we consider this robustness analysis useful for instance for the comparison of the skill profiles of women and men, where the relatively flatter age-skill profile for women observed in Panel A of Figure 2.7 may potentially result from less profile updating among women.

Roughly one third of the workers in our baseline sample (3,238,071 of 8,850,314 individuals) report starting a job at a new employer since 2017. Compared to the baseline sample, the movers are considerably younger on average (34.4 vs. 37.5 years), while the fraction of males is quite similar in both samples (53.2 vs. 54.5 percent).

Panel A of Figure B1 shows the age-skill profile from the baseline sample (in yellow), which is replicated from Figure 2.1, and the corresponding profile for the sample of recent firm switchers (in blue). The age-skill profile of recent firm switchers is notably steeper than for the baseline sample, consistent with switchers being either positively selected on skills, or caring more to update their skill profiles. Still, the qualitative pattern of a concave age-skill profile is just as evident in the sample of recent firm switchers, indicating that this shape is unlikely to be driven by underreporting by inactive users at higher ages.

⁴⁵Results are similar in the smaller sample of those who changed firms within the last year.

Figure B1: Number of reported skills and skill composition by age: Baseline sample vs. recent firm switchers



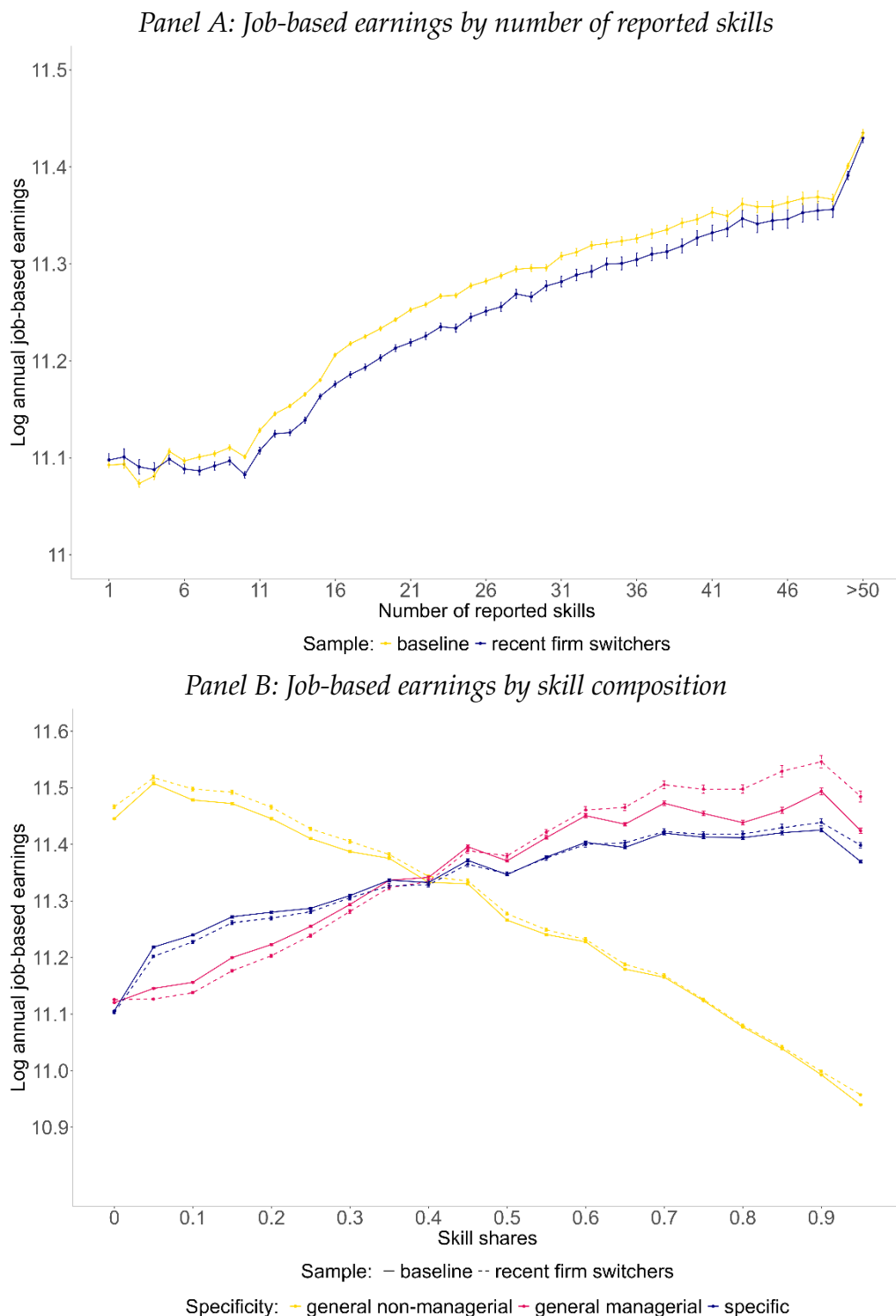
Notes: Average number of reported skills and average fractions of general non-managerial, general managerial, and occupation-specific skills, respectively, by age. Two-year age bins; last bin combines ages 61-64. Error bars depict 99 percent confidence intervals.

Panel B of Figure B1 shows the share of general, managerial, and specific skills by age, as in Figure 2.3, for the baseline sample and for recent firm switchers. The skill composition by age is remarkably similar between the two samples. Likewise, Figure B2 shows that the associations of job-based earnings with the number of reported skills (Panel A) and with the skill composition (Panel B) among workers who have recently changed their employer are very similar to those shown for the baseline sample in Figure 2.5.

Figure B3 reproduces Panel A of Figure 2.8, which reports separate age-skill profiles by gender in both samples. As in the baseline sample, there is only a very small gender gap in the number of reported skills among recent firm switchers in the 20s. But during the 30s and 40s, the gender skill gap grows also among recent job switchers and attains a magnitude that is comparable to the baseline sample. The increasing gender differences in the number of reported skills with age in the baseline analysis are thus unlikely to be driven by gender differences in job mobility.

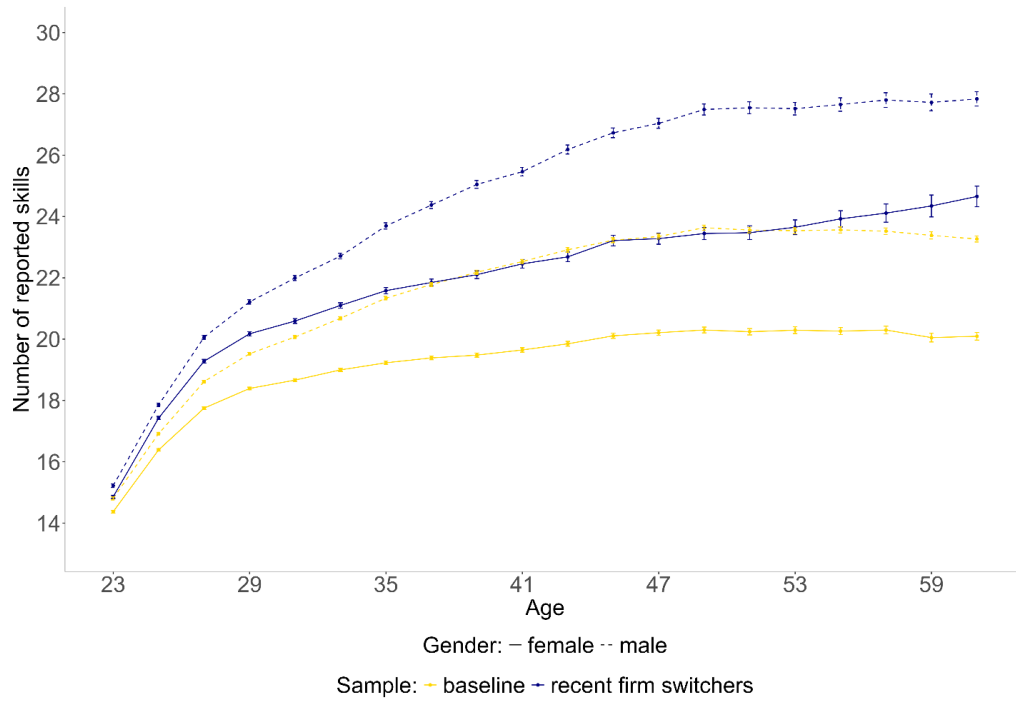
Overall, these key results of our analysis are qualitatively similar for the sample of recent firm switchers, suggesting that the main characteristics of the baseline sample are unlikely to be driven by incomplete updating of inactive users.

Figure B2: Job-based earnings by number of reported skills and skill composition: Baseline sample vs. recent firm switchers



Notes: Average log annual earnings imputed by Revelio Labs’ proprietary salary model by average number of reported skills and average fractions of general non-managerial, general managerial, and occupation-specific skills (five-percent bins), respectively. Error bars depict 99 percent confidence intervals.

Figure B3: Number of reported skills by age and gender: Baseline sample vs. recent firm switchers



Notes: Average number of reported skills by age and gender. Two-year age bins; last bin combines ages 61-64. Error bars depict 99 percent confidence intervals.

2.C Appendix C: Results with ACS Population Weights

Our baseline results give equal weight to each observation in the analysis sample and are thus representative of U.S. college graduates with sufficiently complete LinkedIn profiles. As discussed in section 2.4.4, the composition of LinkedIn users is quite similar but not equal to the composition of the college-educated workforce in the US.

As an alternative to the baseline analysis, we re-estimate key results of our analysis using weights calculated from the pooled 2018-2019 American Community Survey (ACS) data that make the observations of our sample representative in terms of observables for the U.S. college-educated population at large. Both in the LinkedIn and in the ACS data, we compute the distribution of college graduates across 1,536 worker cells delineated by gender, eight five-year age bins (last age bin 58-64), two educational degrees (undergraduate, postgraduate), four U.S. census regions (Midwest, Northeast, South, West), and twelve broad occupational groups (following Deming and Noray 2020).⁴⁶ For the reweighted analysis, all observations of a given worker cell are weighted by the fraction of the cell's share in the ACS vs. in the LinkedIn data.

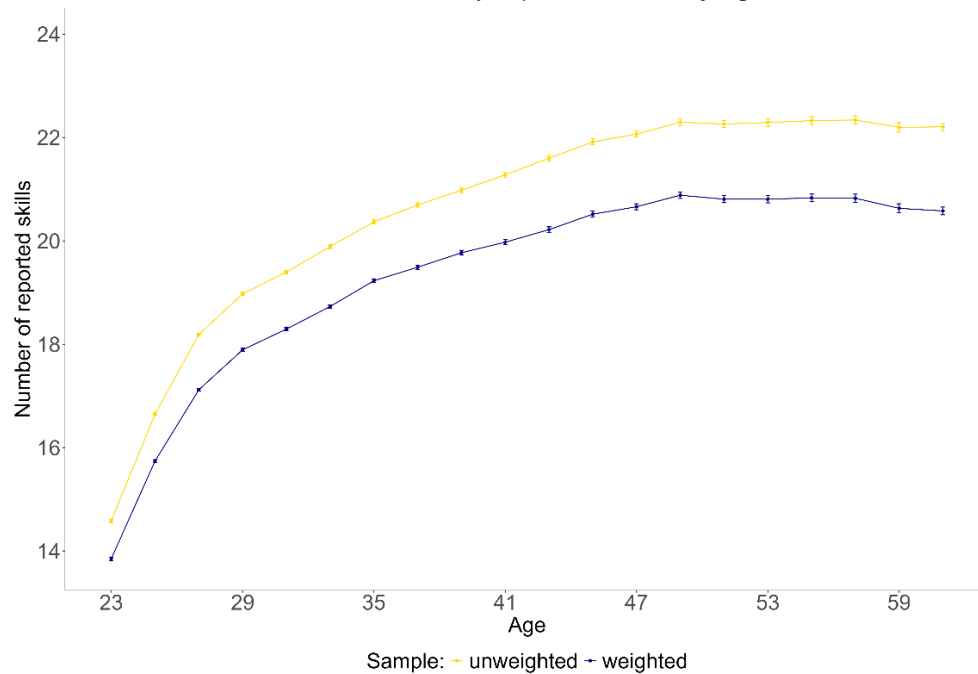
Panel A of Figure C1 indicates that the concave shape of the age-skill profile in Figure 2.1 similarly shows when reweighting the baseline sample to observationally look like the entire U.S. college-educated workforce. Weighting however shifts the graph downwards, indicating positive selection into the analysis sample in terms of number of reported skills. Panel B of Figure C1 shows that reweighting increases the fraction of general skills and reduces the fraction of managerial skills, though the changes are very modest in magnitude.

Panel A of Figure C2 indicates that the positive relationship between number of skills and job-based earnings of Panel A of Figure 2.5 is largely unchanged by reweighting, although there is a level shift to somewhat lower earnings at each level of the skill count. Panel B of Figure C2 confirms the finding of Panel B of Figure 2.5 that job-based earnings are increasing with the fractions of specific and managerial skills and decreasing with the fraction of general skills, and again indicates a lower level of earnings for all skill combinations.

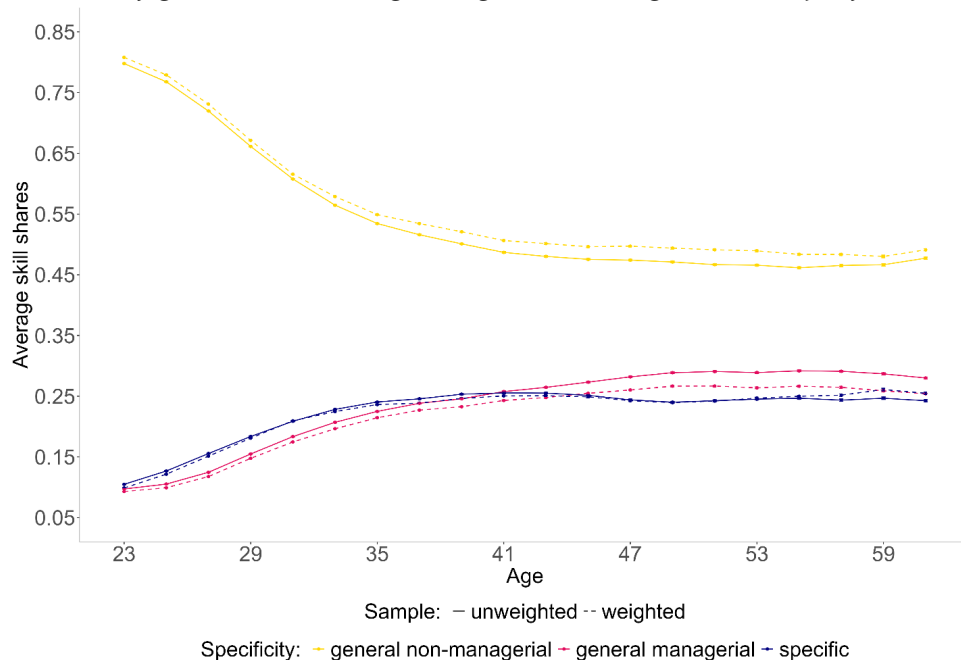
⁴⁶The reweighting analysis omits farming/fishery occupations and military occupations for which cell sizes are very small.

Figure C1: Number of reported skills and skill composition by age: Unweighted vs. weighted data

Panel A: Number of reported skills by age



Panel B: Share of general non-managerial, general managerial, and specific skills by age



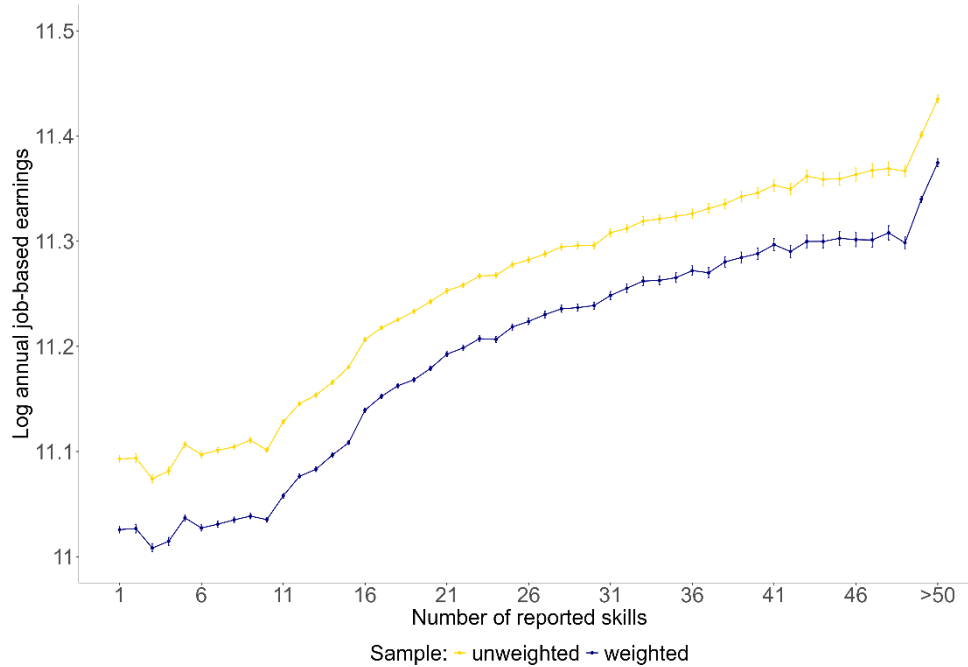
Notes: Average number of reported skills and average fractions of general non-managerial, general managerial, and occupation-specific skills, respectively, by age. Weighted graphs use sampling weights from the 2018-2019 American Community Survey (ACS) to make the estimation sample representative in terms of gender, five-year age bins, educational degrees, U.S. census regions, and occupation cells. Two-year age bins; last bin combines ages 61-64. Error bars depict 99 percent confidence intervals.

Figure C3 shows that the gender-specific age-skill profiles of Figure 2.7 change little with reweighting, except that the level of reported skills is somewhat lower for both women and men at all ages.

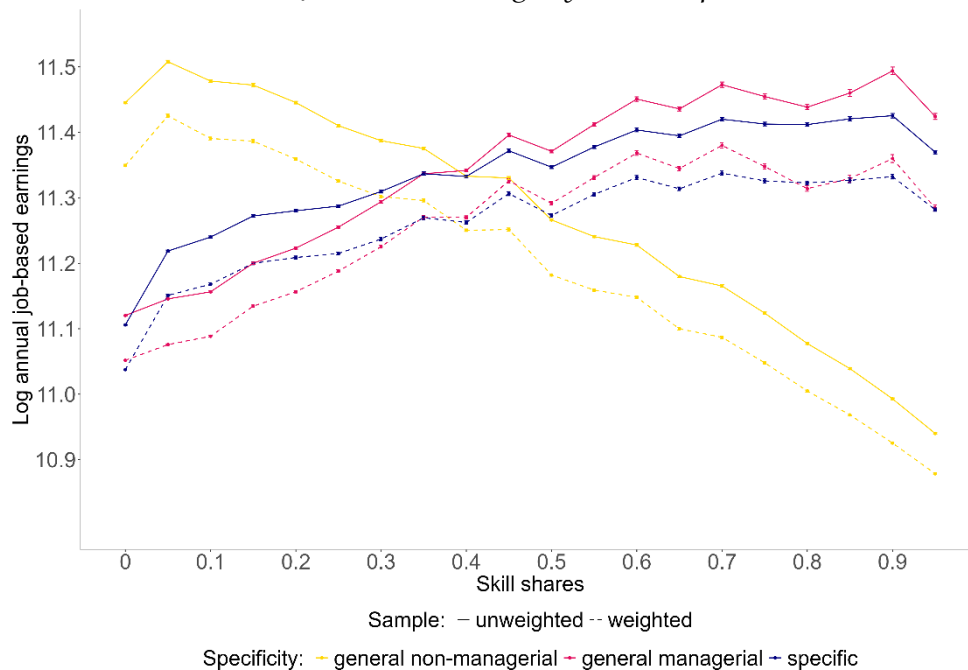
Overall, reweighting the analysis sample to make it demographically representative of the college-educated workforce in the US reveals that our baseline sample is slightly biased towards individuals with higher skill and earnings levels. However, the basic relationships between skills, age, earnings, and gender remain qualitatively and often quantitatively unchanged. The reweighting analysis thus shows that our key findings are unlikely to be driven by patterns of differential participation on the LinkedIn platform for different demographic groups.

Figure C2: Job-based earnings by number of reported skills and skill composition: Unweighted vs. weighted data

Panel A: Job-based earnings by number of reported skills

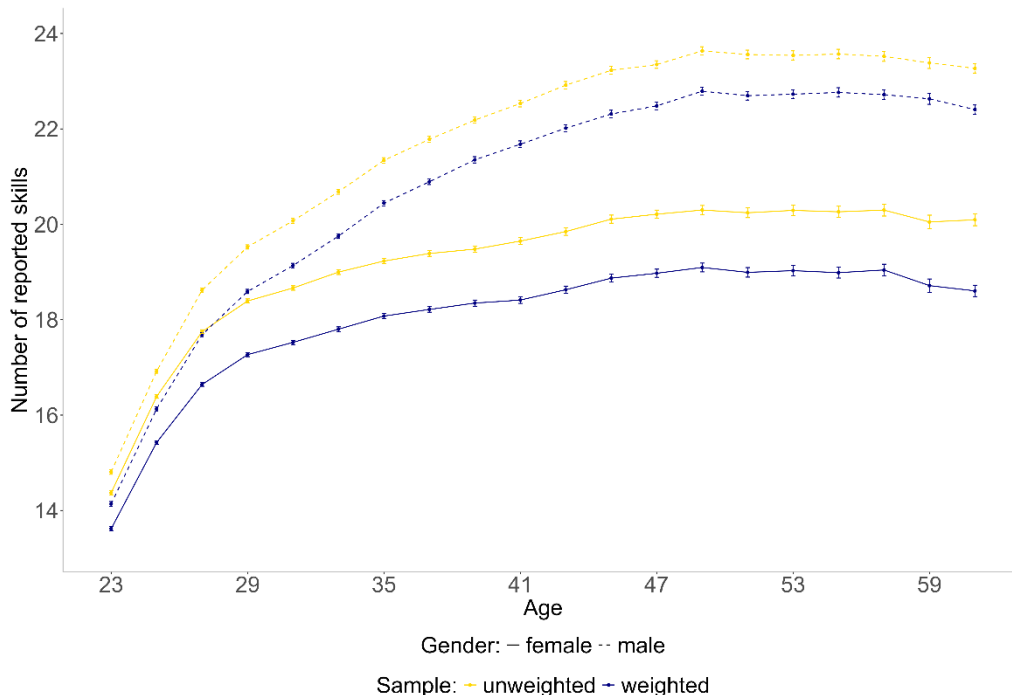


Panel B: Job-based earnings by skill composition



Notes: Average log annual earnings imputed by Revelio Labs' proprietary salary model by average number of reported skills and average fractions of general non-managerial, general managerial, and occupation-specific skills (five-percent bins), respectively. Weighted graphs use sampling weights from the 2018-2019 American Community Survey (ACS) to make the estimation sample representative in terms of gender, five-year age bins, educational degrees, U.S. census regions, and occupation cells. Error bars depict 99 percent confidence intervals.

Figure C3: Number of reported skills by age and gender: Unweighted vs. weighted data



Notes: Average number of reported skills by age and gender. Weighted graphs use sampling weights from the 2018-2019 American Community Survey (ACS) to make the estimation sample representative in terms of gender, five-year age bins, educational degrees, U.S. census regions, and occupation cells. Two-year age bins; last bin combines ages 61-64. Error bars depict 99 percent confidence intervals.

3

Grading Student Behavior¹

3.1 Introduction

Although comportment grading is used worldwide, there is no evidence as to its impact on student outcomes. The policy involves assigning a grade to students' social and work behavior in school and is commonplace in numerous European countries, including Italy, Germany, Poland, and Norway (see Table A1 for an overview). Countries outside Europe such as Japan and Hong Kong follow similar practices, requiring teachers to rate students' behavior on school report cards (Urabe 2006; Cheung and Llu 2000). Comportment grading was also a mainstay in US schools (Maynard 1977; Currie 2004) until their shift towards objective measures of educational output and standards-based grading (Tyre 2010; Duckworth et al. 2012).² Receiving feedback on their comportment in the classroom might encourage students to behave better. However, unlike the "No Excuses" approach adopted by some US charter schools (Angrist et al. 2013), the comportment grading policy focuses solely on grades. Still, it may empower teachers to address disruptive in-class behaviors in a manner that is also noticeable to parents.

Empirical evidence on the causal effects of comportment grading does not exist, and the theoretical case for comportment grading is ambiguous. The merits of

¹This chapter is based on the paper with the same name published in *Labour Economics*, Volume 90, 2024, and co-authored with Lukas Mergele and Larissa Zierow. See Schoner et al. (2024).

²Figures A1 and A2 in the appendix provide famous examples of report cards including comportment grading (also referred to as "deportment" or "conduct" grades) from former US presidents.

comportment grading are thus often debated based on gut-feelings rather than on empirical evidence. Proponents argue that the threat of receiving poor comportment grades might incentivize students to behave better in class and also help them with their transition into the labor market to signal their social skills to employers (Protsch and Solga 2015). Opponents of comportment grades point out that these grades are highly context-dependent, rendering them hardly comparable. This lack of standardization could lead students to feeling they are being treated unfairly when receiving them, which might have demotivating effects on their learning as well as their behavior (Close 2009).

This paper estimates the causal effect of comportment grading on student outcomes exploiting a sequence of reforms across German federal states that introduced comportment grades in schools between 2001 and 2007 as a natural experiment. Due to its federal structure, the German setting offers a rare laboratory to examine education policies within a common political and economic framework. We exploit this policy variation using a staggered difference-in-differences design. After providing evidence that the main identifying assumption – parallel trends – is likely to hold, we run two-way fixed effects (TWFE) regressions. To avoid its pitfalls arising in the presence of heterogeneous or dynamic treatment effects (Goodman-Bacon 2021; de Chaisemartin and D’Haultfoeuille 2020), we adopt the estimation routine put forward in Callaway and Sant’Anna (2021). We establish that heterogeneity in treatment effects are unlikely to matter in our application.

Our primary objective in this study is to investigate the effect of comportment grading on school leavers’ employment or training status. To this end, we use household survey data from the German Mikrozensus. Our results indicate that the effect of comportment grading on the probability of being employed or in training after school is not distinguishable from zero and rather precisely estimated. To better understand the mechanisms behind this outcome, we also analyze students’ non-cognitive skills and reading skills as two channels through which an effect on the school-to-work transitions could operate. Both analyses using representative household surveys and standardized student assessments yield a concordant picture in the sense that point estimates of the reform effect are close to zero. To put our results into perspective, we contextualize potential effect sizes using estimated confidence intervals and the classification proposed by Kraft 2020 and Kraft 2023. Kraft 2020 derives thresholds for effect sizes by assessing the impact of pre-K–12 education interventions on student achievement: Small, i.e., less than 0.05 standard

deviations (SDs), Medium (0.05 to less than 0.2 standard deviations), and Large (greater than 0.2 standard deviations).³ Notably, our analysis indicates that the largest effects (in absolute value) on students' school-to-work transition and non-cognitive skills that we cannot reject fall significantly below the threshold for medium effect sizes (i.e., less than 0.2 standard deviations). However, in the case of reading skills, we cannot dismiss the possibility of observing large effect sizes (i.e., greater than 0.2 standard deviations) considering the wider confidence intervals compared to those of the other outcomes.

Investigating potential explanations for our results, we draw upon three further sources of data. First, we conduct descriptive analyses using cross-sectional report card data on students' comportment and subject grades as well as measures of their non-cognitive skills. We find that the informational value-added of comportment grades about students' non-cognitive skills is limited once subject grades are known. It is therefore unlikely that comportment grades have an impact on students' school-to-work transition through their ability to signal significant non-cognitive skills. Second, we draw upon cross-sectional variation in the PISA data to show that comportment grading is not linked to a more conducive classroom environment for learning. Better classroom behavior usually leads to higher academic achievement, but our finding shows that comportment grades do not correlate with classroom behavior. Hence, it is unlikely that they have an impact on reading skills through this channel. Third, we ran a survey among teachers in Germany. The results suggest that teachers substitute comportment grades with alternative pedagogical disciplinary measures in the absence of comportment grading. Teachers therefore can maintain classroom discipline and students still receive feedback on their behavior regardless of whether comportment grades are provided. This finding offers a possible explanation for a lack of effects on students' skill formation.

Our paper contributes to three strands of literature that investigate inputs to producing student outcomes in the framework of an education production function (e.g., Hanushek 2020; Woessmann 2016). First, we advance the knowledge on the factors within the schooling environment that facilitate a successful school-to-work transition (Ryan 2001). While there is a large literature studying these factors (e.g., Zimmermann et al. 2013), we are the first to focus on the effect of grading students' behavior in school, a widely implemented policy. The

³These benchmarks stem from an analysis of 1,942 effect sizes derived from 747 Randomized Controlled Trials (RCTs).

closest paper to ours in this domain is Protsch and Solga (2015). They use a correspondence study design and find that behavioral reports of students are even more important than GPA for callback rates in the first round of the application process in the German apprenticeship market. However, their contribution is limited to investigating whether there is a direct signalling effect of comportment grades. By contrast, our analysis considers students' academic achievement and non-cognitive skill formation, both of which can have effects on the school-to-work transition themselves.

There are several papers studying the relationship between grades and labor market outcomes. A higher GPA in tertiary education is generally found to increase earnings both in a variety of empirical settings and correspondence studies (Jones and Jackson 1990; Freier et al. 2015; Feng and Graetz 2017; Piopiunik et al. 2020; Tan 2022). Recently, Hansen et al. (2023) have shown that the GPA-related earnings premium quickly fades after labor market entry. By contrast, the counterfactual scenarios in Facchinello (2020) are whether or not students receive grades at all, and therefore more similar to our setting. The author examines a Swedish reform that postponed the introduction of grades in school by several years. He finds that while students from advantaged backgrounds are more likely to be unemployed early in their career, disadvantaged students see their incomes increase. In contrast to that literature, we provide evidence on grades meant to measure behavior, not academic achievement.

Second, we contribute to the understanding of non-cognitive skill formation in school (Bowles and Gintis 2002) by investigating whether they can be fostered through grading comportment. Research on skill formation (e.g., Cunha and Heckman 2007) shows that these skills are malleable, for instance through mentoring programs in childhood and adolescence (Kautz et al. 2014; Kosse et al. 2020; Resnjanskij et al. 2024). There is also evidence for a robust link between these abilities and labor market outcomes (Heckman et al. 2006a; Almlund et al. 2011; Deming 2017; Deming and Kahn 2018), suggesting that an effect on the school-to-work transition could be mediated through improved non-cognitive skills. We are able to test whether students indeed adopt behaviors that are more compliant with conduct requirements in order to obtain positive feedback – as is often proposed as an argument in favor of comportment grades. In Germany, the requirements for good comportment grades include being companionable, diligent, and honest (see

Tables A3 and A2 for more criteria). These concepts are closely related to the non-cognitive skills agreeableness and conscientiousness from the Big Five personality factors, and to trust, which measures pro-social beliefs (Becker et al. 2012).

Third, we investigate whether comportment grades foster performance on standardized reading tests. Literacy skills are associated with better labor market outcomes (e.g., Hanushek et al. 2015) and thus might mediate an effect on students' school-to-work transition. Comportment grades might increase academic achievement by enabling teachers to sanction disruptive behaviors, potentially incentivizing students to behave better in class.⁴ Since less disruptive classrooms enhance academic achievement, comportment grading might benefit human capital formation (Lazear 2001; Angrist et al. 2013; Kristoffersen et al. 2015; Dobbie and Fryer 2020).

The remainder of this paper proceeds as follows: Section 4.2 details the institutional background underlying our work and presents theoretical considerations. Section 4.3 introduces the data sources we use. Section 3.4 outlines how we identify and estimate the causal effect of comportment grading. Section 4.4 presents our results and robustness checks. Section 3.6 offers potential explanations for our findings, Section 3.7 presents an analysis of the costs comportment grades cause, and Section 4.5 concludes.

3.2 Institutional Setting and Theoretical Considerations

3.2.1 Institutional Setting

In Germany, each of the country's 16 federal states is solely responsible for its respective school system. This leads to policy differences across states, although the general structure remains similar. Figure A3 in the appendix provides a graphical overview of the school system. After four years in primary school, children are placed into one of three secondary school tracks: basic school (*Hauptschule*), middle school (*Realschule*), and academic track school (*Gymnasium*). While academic track schools prepare students for studying at university, the other two tracks prepare students for entering the labor market through vocational training.

⁴If comportment grades indeed measure behavior, there could be another way how they enhance academic achievement: Ferman and Fontes (2022) show that teachers inflate subject grades of well-behaved students.

Evaluation of students' comportment is a common practice in these various school types, where students' biannual report card includes grades on working habits and social behavior in addition to subject-specific grades.⁵ Yet, comportment grades have not always been uniformly used across all states and school types. When grading work and social behavior, teachers typically consider students' camaraderie or willingness to cooperate, diligence or work effort, orderliness, and honesty.⁶ Underscoring their significance, these grades are referred to as "head grades" (in German: "Kopfnoten") since they are placed at the top of the report card, above the subject grades. The way in which these grades are presented in the report card can vary from state to state, however. As described in Helbig and Nikolai (2015), some states utilize a five- or six-point scale similar to the normal grading system, while others make standardized written statements about the work and social behavior of students. It is worth noting that the legal basis for grading in the German federal states is primarily criterion-referenced, with individual or collective reference norms rarely used. Under this system, grades reflect the extent to which a student has met the defined requirements, rather than how they compare to their peers or their own previous performance over a certain period of time (Kostorz 2016). As a result, a grading on the curve approach is generally not provided for by the legal framework in Germany, and it is unlikely to be applied when teachers assign comportment grades.

Comportment grades do not determine which secondary school track a student is able to attend. For school-to-work transitions, however, comportment grades signal important non-cognitive skills. Correspondence studies show that comportment grades are an important selection criteria in the apprenticeship market, the main labor market entrance for students without tertiary-level education. Protsch and Solga (2015) demonstrate that employers may value comportment grades even more than regular subject grades.

Comportment grades have a long-standing tradition in German schools, dating back to the inclusion of such grades in scholarship certificates for underprivileged families. Following the Second World War, all German states had incorporated

⁵Comportment grades are typically not utilized during the last two years of schooling in academic track schools. An exception is the regulation of North Rhine-Westphalia after 2007 where comportment grades were mandatory even in the final years of academic track schools. However, this regulation falls outside of our sample period.

⁶As an example, Tables A2 and A3 in the appendix present teacher guidelines for the assessment of behavior in the states of Baden-Wuerttemberg and Saxony.

comportment grades into their curriculum (Arnold and Vollstädt 2001). In the 1970s, public debates on the potential effects of comportment grading in schools led some West German states to dismiss it (Helbig and Nikolai 2015). Prior to reunification, comportment grades were the norm in East Germany, but they were later abolished in several states. We examine the second wave of reforms in East and West German states, which reintroduced comportment grading in the early 2000s. By 2007, all German states had reintroduced comportment grades.

One factor that led to the reintroduction of comportment grades was the increasing pressure from companies and organizations. For instance, Bremen, one of the states examined in this study, implemented comportment grades in response to the success of a project conducted in schools across the state with the help of companies such as DaimlerChrysler.⁷ Additionally, North Rhine-Westphalia implemented comportment grades as a recognition of the increasing importance of "soft skills" in the professional environment.

Figure 3.1 illustrates the adoption of comportment grading in four federal states, namely Bremen (introduced in 2001), Brandenburg (2001), Saxony-Anhalt (2003), and North Rhine-Westphalia (2007), during the period of study (1996 to 2007). It is noteworthy that this policy was implemented by governments of different political affiliations, including center-left and center-right ones, which suggests a non-partisan approach to the issue.

Furthermore, at the federal level, the trend towards reintroducing comportment grades is evident. In 2002, the German parliament established a commission of inquiry on the future of civic engagement, which emphasized the significance of fostering civic engagement in childhood and adolescence. In this context, the use of report cards as a means of recognizing and promoting essential civic traits was explicitly recommended (Enquete-Kommission „Zukunft des Bürgerschaftlichen Engagements“ des Deutschen Bundestages 2002).

⁷The project involved the development and evaluation of appropriate assessment methods for working habits and social behavior over a period of more than two years. The implementation of the project was based on various factors, including the results of workshops and surveys. Additionally, the regulation was informed by associated school research projects on key qualifications and assessment methods, as well as the dialogue with representatives from the business sector, and the example of corresponding forms of assessment of work and social behavior in the training system of DaimlerChrysler. The regulation also took into account similar regulations and documented experiences in other federal states, especially in Lower Saxony, Brandenburg, and Thuringia, as well as the evaluation of relevant assessment methods used in primary schools and comprehensive schools in Bremen. Finally, the proposed regulation underwent a "trial evaluation" of the presented drafts. Information available at: <https://www.bildung.bremen.de/sixcms/media.php/13/1129.pdf>, last access on March 12, 2023.

Despite these developments, the issue of comportment grading remains contentious. After introducing this grading system in 2007, the new center-left coalition in North Rhine-Westphalia abolished it again in December 2010, following public demonstrations against it. Conversely, in 2013, Mecklenburg–Western Pomerania switched from a written assessment to a classic German grading system using a six-point scale ranging from 1 to 6 for comportment grades, while also making it mandatory for primary school children. In Bavaria, comportment grades in primary schools became compulsory from 2008 onwards.

3.2.2 Theoretical Considerations

From a theoretical perspective, comportment grading in school could affect student outcomes via the following channels.

(i) If students receive feedback on non-cognitive dimensions of their skills, they may be motivated to invest into these skills in order to get more positive feedback. This would be in line with literature showing that grades, in general, can serve as incentives in the schooling context and that these incentives are important for students' educational investments (Hvidman and Sievertsen 2021). Being companionable, diligent, and honest, for example, are among the criteria teachers ought to consider when grading comportment in Germany (see Table A2). At the same time, they are closely related to the non-cognitive skills agreeableness and conscientiousness from the "Big Five" personality factors, and trust, which measures prosocial beliefs (Becker et al. 2012).

Research has shown that these non-cognitive skills can be easily influenced and shaped (e.g. through mentoring programs) in childhood and adolescence (Kautz et al. 2014; Kosse et al. 2020; Resnjanskij et al. 2024) and that these skills and later labor market outcomes are strongly linked (Heckman et al. 2006a; Almlund et al. 2011; Deming 2017; Deming and Kahn 2018). Consequently, if students indeed invested more into their non-cognitive skills when they receive comportment grades, we would expect more favorable labor market outcomes, i.e. in our setting, a more successful transition into the labor market after school. Furthermore, we would expect to see an increase in non-cognitive skill measures.

(ii) If students receive feedback on their comportment in the classroom, this might have a disciplinary effect. In contrast to the "No Excuses" approach (e.g., Angrist et al. 2013) adopted by charter schools in the US, the comportment grading policy stands

in isolation, that is, it does not affect the school environment beyond grades. Still, it enables teachers to sanction disruptive in-class behaviors in a way that is also visible to parents. If classroom discipline improves due to this rather unobtrusive policy, something teachers appear to affirm (see Table 3.10), this would have beneficial repercussions on human capital formation (Lazear 2001). Research has shown that disruptive classroom behavior decreases student achievement significantly (Kristoffersen et al. 2015; Ahn and Trogdon 2017). Therefore, comportment grading might lead to better reading skills, which would also lead to a higher probability to enter the labor market successfully after school.

(iii) To the extent that comportment grades can also serve as indicators of non-cognitive skills (Landersø and Heckman 2017), as supported by respondents in our teacher survey (see Figure 3.3 and Table 3.10), they might enable students to signal these skills to employers, thereby reducing information asymmetry and facilitating students' transition into the labor market (Protsch and Solga 2015).

(iv) Yet critics of comportment grades argue that the presence of different settings when establishing these grades renders them hardly comparable.⁸ Due to a lack of standardization, students might question the fairness behind the grading process, which could in turn have demotivating effects on their learning as well as behavior (Close 2009). Furthermore, intrinsic motivation for good behavior could be crowded out by grade-driven extrinsic motivation (see Koch et al. 2015, for a comprehensive discussion of such motivational crowding-out in the context of educational interventions). In this scenario, we would expect negative effects on student outcomes.

3.3 Data

To investigate whether the introduction of comportment grading affects students' school-to-work transition and skill formation, one would ideally draw on a single set of panel data with detailed information on individuals' schooling history, skill measures, and employment records. Given the lack of such a dataset, we compile repeated cross-section data drawn from three different sources. First, we use household survey data – the German Mikrozensus – to get information about individuals' school-to-work transition. Second, we use individual-level survey

⁸This is contested by teachers: 77% of participants in our survey state that they determine comportment grades based on official criteria (see Figure 3.3).

measures of respondents' non-cognitive skills from the German Socio-Economic Panel. Third, data on ninth-grade literacy test scores are drawn from student assessment studies. Applying the same set of sample restrictions across datasets makes this data well-suited to test our hypotheses. All of the data sources provide individuals' year of birth and their federal state of schooling or federal state of residence at the time the survey was conducted. We impute their year of enrolment and federal state of schooling from this information. Then, we link the individual date via the year-of-enrolment and state identifier with the reform status of their state of schooling. By doing so, individuals are assigned as being affected by the reform (treated) if the reform was in place when they entered school.⁹ For these three sources, we add state-level information about whether schools grade compartmentment to derive treatment and control group assignments.

To explore potential explanations for our findings, we draw on two further data sources. Cross-sectional report card data is taken from the National Educational Panel Study (NEPS) to investigate the relationship between subject and compartmentment grades, and make sense of our reform analysis. Finally, we ran a survey among teachers in Germany eliciting their assessment of compartmentment grades.

Data on the school-to-work transition The German Mikrozensus is a household survey covering one percent of the German population in annual waves since 1970. An appealing feature of the data is that participation is mandatory, which is why unit non-response rates are consistently below 4% during the time frame we consider (Länder 2018). We make use of the 2011–2018 waves and restrict the sample to individuals aged between 15 and 20¹⁰ living in a state that introduced compartmentment grading. We exclude individuals who still attend secondary school and those studying towards a university entrance degree who have yet to transition into the labor market. Thus, we focus on students who have completed secondary

⁹Information on the federal state of schooling is available in the nation student assessment data as well as partially in the SOEP data. However, the Mikrozensus lacks this information. We use the SOEP data and information from the German statistical office to gauge what share of individuals is assigned the wrong federal state of schooling if one instead uses current federal state at ages 15 to 20. We thereby achieve misclassification rates of 2% for the SOEP and 4.2% for the Mikrozensus data. See Appendix 3.E for more detailed information.

¹⁰Pinquart et al. (2003) argue that one should consider up to five years after regular school-leaving age for assessing students' school-to-work transition. Regular school-leaving age for the school tracks we consider is 15 and 16 for the low and medium track, respectively (see Figure A3).

education, which enables them to start working directly after school or to begin vocational training (“Ausbildung”).¹¹

We use information on individuals’ employment status to derive a binary measure capturing successful school-to-work transitions. More specifically, we consider an individual to have successfully transitioned from school to work if they are in vocational training, completing a secondary-schooling degree after finishing a lower one, or is employed at least part-time. Conversely, unsuccessful transitions include individuals that are marginally employed, looking for work, or temporarily out of the labor force.¹²

Table 3.1: Descriptive statistics Mikrozensus

	Mean	Std. Dev.	Min	Max	Observations
Successful School-to-work Transition	0.86	0.34	0	1	16982
Successful School-to-work Transition (strict)	0.80	0.40	0	1	16982
Comportment Grading	0.09	0.29	0	1	16982
Female	0.42	0.49	0	1	16982
Migrant	0.30	0.46	0	1	16982

Notes: Sample includes students from the federal states of Bremen, Brandenburg, Saxony-Anhalt, North Rhine-Westphalia. “Comportment Grading” is an indicator variable equaling 1 if comportment grading was in place when the individual was enrolled in school. “Successful School-to-work Transition (strict)” is an alternative measure of successful school-to-work transition, excluding employed individuals who have not earned any vocational qualification prior to their employment.

Sources: Mikrozensus waves 2011–2018.

Table 3.1 provides the descriptive statistics of this sample, which consists of 16,982 observations, covering the reforming states of Bremen, Brandenburg, Saxony-Anhalt, and as non-reforming control state North Rhine-Westphalia.¹³ The mean of our main outcome variable “Success” is quite high: 86% of the observed individuals are in vocational training or catching up on a secondary-schooling degree or are employed at least part-time. The indicator variable “Comportment Grading”

¹¹To avoid potential sample selection issues caused by changes in student tracking resulting from the reform, we investigate whether the reform had any effect on tracking (see Table A9). Fortunately, our analysis does not find any significant impact of the reform on tracking. As a result, we are confident that our estimation approach is not affected by sample selection issues.

¹²This corresponds closely to the definition of “out-of-school joblessness” given in Ryan (2001), which includes those unemployed according to the ILO/OECD definition and those not enrolled in an educational course. We add the marginally employed to this group since we are interested in transitions from school into stable employment relationships.

¹³See Section 3.4 for our discussion of the definition of treatment and control groups in this setting

shows that 9% of the individuals had comporment grading in place when they were enrolled in school.

Non-cognitive skill measures Survey measures of non-cognitive skills are taken from the German Socio-Economic Panel (SOEP, see Goebel et al. 2019), a survey data set representative of private households in Germany. From the SOEP, we build a cross-section of individuals aged 15 to 20 from different survey years (2003 – 2020) born between 1990 and 2000.

We investigate the formation of non-cognitive skills that directly relate to criteria teachers are expected to consider when grading comporment (Tables A2 and A3). We focus on agreeableness and conscientiousness from the “Big Five” personality factors, which overlap with the “Camaraderie” and “Work effort” or “Diligence” criteria. We also investigate an individual’s level of trust, which potentially affects one’s willingness to be honest, and is therefore related with the “Honesty” dimension. Each of these latent concepts is measured using answers to three survey items on Likert-type scales. To generate a single measure for each concept, we average the score of each item for each individual. If measures from different survey years are available for a given individual, we take the earliest available measure.

Table 3.2: Descriptive statistics SOEP

	Mean	Std. Dev.	Min	Max	Observations
Trust	-0.00	1.00	-2.51	2.81	5547
Conscientiousness	-0.00	1.00	-3.41	1.80	5547
Agreeableness	0.00	1.00	-4.56	1.71	5547
Comporment Grading	0.64	0.48	0.00	1.00	5547
Female	0.50	0.50	0.00	1.00	5547
Migrant	0.29	0.45	0.00	1.00	5547

Notes: Sample includes students from the federal states of Bremen, Brandenburg, Saxony-Anhalt, North Rhine-Westphalia, Baden-Wuerttemberg, Berlin, Hamburg, Mecklenburg-Western Pomerania, Rhineland-Palatinate, Saarland, Saxony, and Schleswig-Holstein. "Comporment Grading" is an indicator variable equaling 1 if comporment grading was in place when the individual was enrolled in school.

Sources: SOEP-Core v37.

Table 3.2 provides the descriptive statistics of our SOEP sample, which consists of 5,547 observations, covering the reforming states of Bremen, Brandenburg, Saxony-Anhalt, the as non-reforming control state North Rhine-Westphalia, and the "always-treated" states Baden-Wuerttemberg, Berlin, Hamburg, Mecklenburg-Western

Pomerania, Rhineland-Palatinate, Saarland, Saxony, and Schleswig-Holstein, see Figure 3.1 for an overview of the reform timing. The non-cognitive skill measures are standardized to have mean zero and unit standard deviation. The indicator variable "Comportment Grading" shows that 64% of the individuals had comportment grading in place when they were enrolled in school.

Student assessment data Measures of students' reading skills are taken from the German extension of the Programme for International Student Assessment (PISA-E), which is available with federal state identifiers for the years 2000, 2003, 2006, and 2012. For the years 2009 and 2015, we employ data from the National Assessment Study by the Institute for Educational Quality Improvement (IQB), which is collected in accordance with PISA. The data were made available by the Research Data Centre at the Institute for Educational Quality Improvement (FDZ at IQB). All achievement tests target students in ninth grade and are always performed between May and July. As participating schools within each state are drawn at random, each wave constitutes a cross-section of ninth graders that is representative at the state level. Taken together, these waves form a pseudo-panel of German states from 2000 to 2015, with observations occurring every three years. We impose identical sample restrictions as used for the census data wherever possible.

In addition to compulsory tests that measure students' reading skills, questionnaires are given to schools, students, and parents, which elicit a wide range of socio-demographic background characteristics. Test scores are standardized and comparable across waves. We standardize reading test scores to have mean zero and unit standard deviation. Reading test scores were generated as follows: To keep the length of student achievement tests tractable, test providers estimate individual test scores based on a random subset of the full questionnaire (Jerrim et al. 2017). Hence, five different estimates of reading test scores ('plausible values') are available throughout all waves to represent the distribution of *true* reading skills. We do not consider math skills as they are only tested in every other wave of the National Assessment Study.¹⁴

¹⁴For more details on the data, refer to Baumert et al. (2002), Prenzel et al. (2007), Prenzel et al. (2010), Prenzel et al. (2019), Sachse et al. (2012), and Schipolowski et al. (2019).

Table 3.3: Descriptive statistics nationwide student assessments

	Mean	Std. Dev	Min	Max	Observations
Reading skills	0.01	1.00	-6.00	5.11	128,249
Academic track school attendance	0.33	0.47	0.00	1.00	128,249
Comportment Grading	0.72	0.45	0.00	1.00	128,249
Female	0.49	0.50	0.00	1.00	128,249
First generation migrant	0.12	0.33	0.00	1.00	128,249
Age (months)	187.26	6.39	148.96	242.00	128,249
Low SES	0.30	0.46	0.00	1.00	128,249

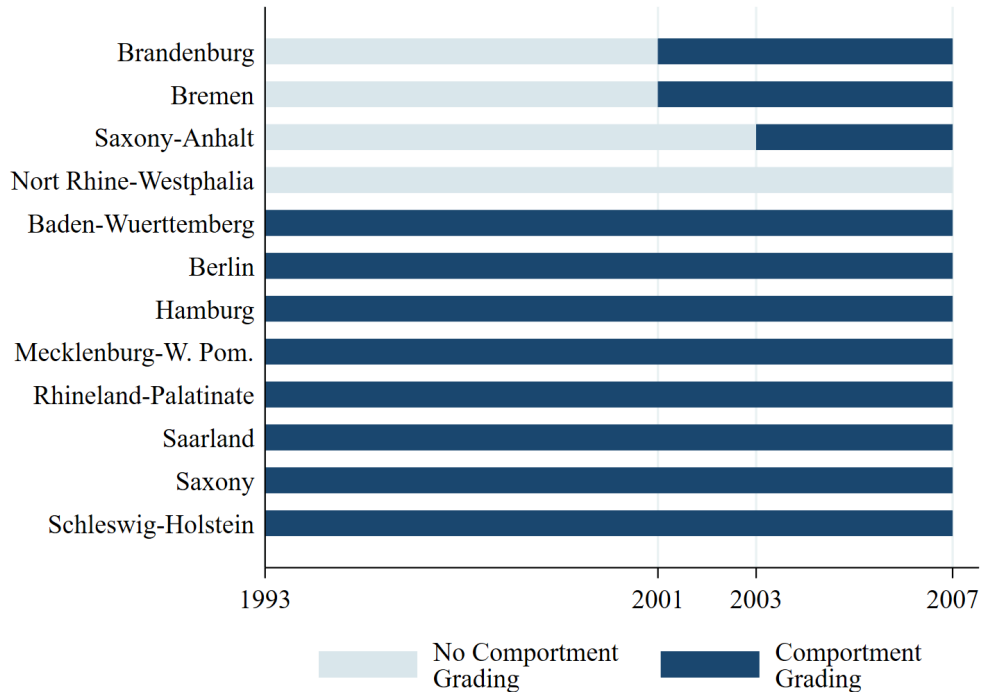
Notes: Sample includes repeated cross-sections of students in ninth grade from the federal states of Bremen, Brandenburg, Saxony-Anhalt, North Rhine-Westphalia, Baden-Wuerttemberg, Berlin, Hamburg, Mecklenburg-W. Pom., Rhineland-Palatinate, Saarland, Saxony, Schleswig-Holstein. "Comportment Grading" is an indicator variable equaling 1 if comportment grading was in place when the individual was enrolled in school. Low SES defined as parents having obtained ISCED Level 3B/C at most.

Sources: PISA 2000, 2003, 2006, and 2012, IQB-LV 2008-9 (v2), IQB-BT 2015 (v5).

Table 3.3 provides the descriptive statistics of our student assessment data, which consists of 128,249 observations covering the reforming states of Bremen, Brandenburg, Saxony-Anhalt, the as non-reforming control state North Rhine-Westphalia, and the "always-treated" states Baden-Wuerttemberg, Berlin, Hamburg, Mecklenburg-Western Pomerania, Rhineland-Palatinate, Saarland, Saxony, and Schleswig-Holstein (refer to Figure 3.1 for an overview of the reform timing). The outcome variable "Reading skills" is based on the PISA reading test scores. The variable "Academic track school attendance" shows whether an individual attends the most demanding track in the German school system, leading to the university entrance degree ("Abitur"), and is used for analyzing whether the composition in German school tracks has changed due to the reform, see Table A9. The indicator variable "Comportment Grading" shows that 72% of the individuals had comportment grading in place when they were enrolled in school. For an exploratory analysis of classroom discipline in Section 3.6, we also use PISA 2000 as a cross section because it includes item batteries on classroom discipline that were separately answered by students and school principals, see Table 3.9

Data on comportment grading reforms Data on state-level comportment grading policies were gathered from school reform coding based on the states' schooling legislation and collected by Helbig and Nikolai (2015). We classify state policies according

Figure 3.1: The introduction of compartment grading over time and by state within the sample



Notes: Compartment grading is defined as the implementation of either mandatory written or mandatory numerical compartment grading. German states that reformed compartment grading shortly ahead of our sample period were excluded from the overview. We also exclude these states from our analysis. North Rhine-Westphalia did not introduce compartment grading until 2007.

Sources: Own representation based on Helbig and Nikolai (2015).

to four categories: (1) no compartment grading, (2) optional written compartment grading, (3) mandatory written compartment grading, or (4) mandatory numerical compartment grading. Given that we are interested in the effect of compartment grading per se, we consider individuals experiencing any kind of mandatory grading of compartment as treated ((3) and (4)) and the others as non-treated ((1) and (2)).

As Figure 3.1 shows, four federal states adopted compartment grading during our sample period (1996 to 2007). More specifically, Bremen and Brandenburg constitute the first treatment group, which introduced compartment grading in 2001; Saxony-Anhalt followed suit in 2003 and therefore serves as the second treatment group. Finally, North Rhine-Westphalia did not adopt compartment grading until 2007 and consequently serves as our control group. The eight other federal states shown already had compartment grading schemes in place prior to our sample

period.¹⁵ After having established that heterogeneity in treatment effects is unlikely to matter in our case, we use them to increase statistical power.

Individuals are considered treated if compartment grading was in place in the year they enrolled in school, i.e., treated students received these grades throughout their school career since none of the reforms were revoked. Based on our analysis of legal documents related to the reform, we chose this assignment method because the definitive treatment was only introduced for new student cohorts.¹⁶ According to the legal rules, we found no evidence that the reform had to be applied retroactively to students who were already enrolled in school. To address any remaining concerns, we conduct a robustness test by assigning individuals to the treatment group if they had already been enrolled in school for some years at the time of the reform's implementation.

Report card data We use the Starting Cohort 3 from the National Educational Panel Study (NEPS SC3, version 11.0.0) to compare actual compartment grades with subject grades included in report cards (Blossfeld and Roßbach 2019). The first wave from fall 2010 includes individual-level data for students in grade five. These individuals were resurveyed at regular intervals until wave 10, which was conducted in fall 2018 when students were about 19 years old. Compartment grades were elicited in waves eight to ten, referring to students' respective final report card at graduation. We also retrieve the final grade-point average (GPA) as well as subject grades in Math and German. To harmonize the grading information, we round all grades to the next integer as reporting formats differ across grades. We also reverse the standard German grading scale to ease interpretation, such that higher numbers indicate better grades. This implies that our grades range from 1 ("insufficient") to 6 ("very good"). Moreover, we use data from wave 10 for information on agreeableness and conscientiousness from the "Big Five" personality factors. We focus on students who received compartment grades within their final report cards. If individuals achieve more than one school degree, we keep the first one that contained compartment grades on the final report card.

¹⁵We cannot include all 16 federal states since the remaining 4 had either introduced or abolished compartment grading shortly before our sample period starts. This could confound our results.

¹⁶In the first and second grades, compartment grading was usually done in writing, while from the third grade onwards, numerical grading became mandatory. See for example here, last access on March 13, 2023.

Table 3.4 provides the descriptive statistics of our grading data, which consists of 886 students and their report cards.

Table 3.4: Descriptive statistics NEPS

	Mean	Std. Dev.	Min	Max	Observations
German Grade	4.33	0.80	2.00	6.00	891
Math Grade	4.17	0.95	2.00	6.00	891
GPA	4.52	0.64	3.00	6.00	891
Conscientiousness	0.006	1.01	-3.15	2.10	891
Agreeableness	-0.006	1.01	-3.17	2.57	891
Comportment Grade	4.97	0.70	1.00	6.00	886

Notes: Sample includes German students that were interviewed in fifth grade. Comportment grades are available in a student's graduation report. The subject grades represent half-year grades and, like the GPA, are taken from the graduation year. Subject grades and GPA are rounded to integers. Non-cognitive skills are standardized. All variables assume that higher values are better.

Sources: NEPS SC3 11.0.0

Teacher survey on comportment grading As a last step, we conduct an expert survey on the specifics of comportment grading among a convenience sample of 246 teachers in Germany. The respondents are recruited by distributing the survey on online teacher platforms, via teacher groups on social media, and by directly contacting schools. Overall, this online survey enables us to gain a more nuanced impression of how comportment grading takes place in practice. We are particularly interested in the teachers' assessment of the effectiveness of comportment grading, the time spent per student and report card as well as the number of teachers typically involved, and teachers' grading standards.

Table A4 compares respondents' demographic characteristics to national averages from official statistics. It provides assurance that our sample – although not representative in nature – does not stray too far from the true distribution of age, gender, place of work, and school type among teachers in Germany.

3.4 Empirical Strategy

Identifying the average effect of comportment grading on the students who received comportment grades relies on the staggered adoption of comportment grading across federal states, which gives rise to a generalized difference-in-differences approach.

Therefore, we are interested in the coefficient δ of the following regression

$$Y_{ist} = \gamma_s + \lambda_t + \delta \cdot CG_{st} + \mathbf{X}_{ist}^\top \boldsymbol{\beta} + \varepsilon_{ist}, \quad (3.1)$$

where Y_{ist} is an outcome for student i attending school in state s in cohort t . CG_{st} is a dummy variable indicating that schools in this state graded compartment for this cohort (from the year of enrollment onwards) and zero otherwise. \mathbf{X}_{ist} contains an individual's sex and migration background to increase precision. Furthermore, we include a set of fixed effects capturing the federal state of schooling (γ_s) and year of enrollment (λ_t). ε_{ist} is an error term.

Standard errors are clustered at the level of treatment assignment, that is, federal states (Abadie et al. 2022). In our main specification, we use 4 and 12 federal states, respectively, of which three introduce compartment grades at different points in time (see Figure 3.1). To account for the small number of clusters as a potential source of bias in the coefficients' variance estimates (e.g. Cameron et al. 2008), we apply the wild cluster bootstrap (WCB) procedure outlined in Roodman et al. (2019) to obtain valid p -values.

Ordinary least squares estimates of δ using the TWFE specification above capture a causal effect only if treatment effects are homogeneous across time and units (de Chaisemartin and D'Haultfœuille 2020; Goodman-Bacon 2021). This is because the TWFE estimator corresponds to a weighted average of all possible 2x2 difference-in-means estimates during the sample period. These include invalid comparisons of newly-treated to already-treated units. If treatment effects evolve over time, the estimated 2x2 effects from invalid comparisons might be weighted negatively, that is, they are subtracted from the estimate when being aggregated to a single measure (Goodman-Bacon 2021). Although dynamics might be a lesser concern here, we want to ensure the robustness of our estimates regarding the issues arising from treatment effect heterogeneity.¹⁷ Therefore, we implement the estimator proposed by Callaway and Sant'Anna (2021), henceforth (C/S), which excludes states that already used compartment grading schemes prior to our sample period. It is robust against both forms of treatment effect heterogeneity and differs from the TWFE approach mainly by ensuring that newly-treated units are only compared to not-yet-treated units. As

¹⁷Since we use repeated cross-section data, dynamic treatment effects would be equivalent to assuming cross-cohort spillover effects. More specifically, treatment effects would need to be a function of the number of cohorts that had already been treated prior to the current cohort. This is because we do not observe individuals repeatedly, that is, there is no way treatment effects can evolve for individuals.

detailed below, the C/S routine also estimates a weighted average of 2x2 effects.¹⁸ Each 2x2 effect is an estimate of the average treatment effect on the treated (ATT):

$$\text{ATT}(g, t) = E(Y_t(g) - Y_t(0) | G = g),$$

where g denotes the year group G first receives the treatment and t are time periods. In our setup, there are two treatment groups receiving treatment in 2001 and 2003, respectively ($g \in \{2001, 2003\}$). We restrict the sample period to $t = 1992, \dots, 2006$ since North Rhine-Westphalia introduces compartment grading in 2007. This means that we have 14 ATTs for each group, 8 (10) pre-treatment and 6 (4) post-treatment effects for the group with $g = 2001$ ($g = 2003$). To obtain a single estimate that can be interpreted as a multi-period and multi-group extension of the ATT in the canonical 2x2 design, we first average over post-treatment effects for each group and then across treatment groups to obtain

$$\begin{aligned} \text{ATT} := & \frac{1}{6} \sum_{t=2001}^{2006} \text{ATT}(g = 2001, t) \cdot \Pr(G = 2001) \\ & + \frac{1}{4} \sum_{t=2003}^{2006} \text{ATT}(g = 2003, t) \cdot \Pr(G = 2003), \end{aligned} \quad (3.2)$$

whose estimates can be compared directly to estimates of ATT obtained from TWFE specifications. If estimates on the same sample differ substantially, there must be heterogeneity in treatment effects either across time, across units, or both.

We run both estimation routines on the exact same set of individuals for our analysis of students' school-to-work transition to investigate whether our results suffer from the negative weighting issue. Since this is not the case, we add further federal states that did not change their compartment grading policy between 1993 and 2007 to our sample for the remainder of our analysis (see Figure 3.1). This increases statistical power and alleviates the inference issues arising from few clusters.

Note that the C/S approach does not allow us to conduct cluster-robust inference due to the small number of clusters (Callaway and Sant'Anna 2021, p.25). We therefore report heteroskedasticity-robust simultaneous confidence bands. In contrast to usual practice, simultaneous instead of pointwise confidence intervals capture the estimation uncertainty arising from estimating the whole sequence

¹⁸See the Appendix 3.B for a more detailed exposition.

the group-time average treatment effects that go into the aggregation to obtain a single estimate of the treatment effect (see Figure 3.2). This automatically implies robustness against multiple hypothesis testing, which would be a concern if we used pointwise confidence intervals since each pre- and post-treatment coefficient in Figure 3.2 corresponds to one hypothesis test.

C/S confidence intervals grow in the number of estimated effects, that is, in the number of time periods. To maximize power, one can reduce the number of estimated effects in the following way. Our setup gives rise to four time periods: before the first group gets treated (1992 to 2000), after the first group got and before the second group gets treated (2000 to 2002), after the second group got treated and before North Rhine-Westphalia adopts the treatment (2003 to 2006), and 2007 and later. Following the same logic as above, this gives 0 (1) pre-treatment and 2 (1) post-treatment effects for the group with $g = 2001$ ($g = 2003$). This setup still allows for pre-testing and, crucially, leads to much narrower confidence intervals (see Figure A5 and Panel B of Table 3.5). Since estimates of the overall ATT rely on averaging of post-treatment effects in Table 3.5 (see Appendix 3.B), point estimates barely change between Panels A and B of Table 3.5.

There are two main threats to identifying the causal effect of compartment grading on treated students' outcomes. First, we need to assume parallel trends. This means that, in absence of the reforms, student outcomes would have followed the same trajectory over time for both treatment groups relative to the respective control group. Although fundamentally untestable, we corroborate this assumption by investigating pre-treatment trends in an event-study specification using our main dataset, the Mikrozensus data.¹⁹ Note that we allow parallel trends to hold only after conditioning on student sex and migration background since the latter is unbalanced across groups and potentially affects the evolution of student outcomes (Abadie 2005; Heckman et al. 1997). Figure 3.2 shows that none of the individual pre-treatment coefficients are statistically distinguishable from zero for both treatment groups, suggesting that the parallel trends assumption is likely to hold.

Another threat to identification arises from different school reforms being introduced at the same time as compartment grading. We investigated the compendium of German school reforms since World War II by Helbig and Nikolai

¹⁹Note that our event-study results are based on the approach put forward in Sun and Abraham (2021) and therefore not biased by treatment effect heterogeneity.

(2015) and did not find any concomitant school reform. The only exception is Saxony-Anhalt, where compartment grading was introduced in parallel to a shortening of the duration of primary school from six to four years. We address this potential concern by analyzing the two treatment groups that introduced compartment grades in different years separately (Figure 3.2), showing that the post-treatment effects have similar sizes in both groups. Another way to see this is by averaging the post-treatment effects for both groups separately, which shows that, reassuringly, both groups experience null effects (Figure A4).

Finally, if there were differences in the selection of students into school tracks over time along with the reforms, this would bias our results. To alleviate these concerns, we estimate the effect of the compartment grading reforms on the likelihood to attend the academic track. Using the student assessment data, we show in Table A9 that there is no significant effect on academic track attendance.

3.5 Results

3.5.1 Compartment Grading and the School-to-Work Transition

Panel A of Table 3.5 displays C/S estimates of the aggregated ATT estimand in equation 3.3 (columns 1 and 2)²⁰ and from estimating equation 3.1 (columns 3 and 4) using TWFE. Even-numbered columns add student sex and migration background as control variables. Point estimates obtained from C/S imply that the compartment reform-induced change in the probability of transitioning from school to work successfully are very close to zero, amounting to 0.28 and -0.22 percentage points in columns 1 and 2, respectively.

Columns 3 and 4 display the results from estimating equation 3.1. Estimated effect sizes hardly differ across estimation techniques, corroborating the notion that heterogeneous treatment effects and ensuing negative weights issues are a lesser concern in our setup. Note that while confidence intervals based on cluster-robust standard errors are smaller than those robust to multiple hypothesis testing, we expect the former to be too narrow given the small number of clusters. Wild cluster bootstrap p -values alleviate this issue and similarly indicate that we fail to reject the hypothesis of a null effect by a wide margin.

²⁰Note that these numbers stem from the same estimation conducted for Figure 3.2. They can be obtained by averaging the post-treatment point estimates across both groups and time periods as detailed in Appendix 3.B.

Panel B reduces the number of time periods as outlined in Section 3.4.²¹ The width of confidence intervals in columns 1 and 2 decreases substantially by more than 30% compared to Panel A, while point estimates remain very close to zero. Confidence intervals corresponding to TWFE estimations barely change since there is no multiple hypothesis testing correction. These results sharpen our inference in the sense that we can reject effect sizes beyond 3.3 percentage points in absolute value. Put differently, non-rejectable effect sizes are larger than one tenth of a standard deviation (the standard deviation of the variable successful school-to-work transition is 0.34; see Table 3.1).

For consistency with the remainder of our analyses, we also report TWFE estimates for the full sample of 12 federal states instead of solely using the four states in Table 3.5. The results are shown in Table 3.6. Although point estimates are not as close to zero as in our other specifications, we still do not reject the null hypothesis.

Robustness checks To alleviate concerns about the exact treatment timing as explained in Section 4.2, we report results using delayed treatment assignment. More specifically, we define the treatment as applying to all students who were in second, third, or fourth grade and higher instead of first grade, respectively, at the time of the reform. Appendix Table A6 shows that our null result persists in Panels B and D. If we define treatment assignment as applying to third graders and above in Panel C, we find a negative effect that is borderline significant. Yet, considering the few-cluster adjusted inference from the WCB, we do not reject the null.

Further robustness checks in Appendix Table A5 use different definitions of our outcome variable and sample restrictions. Panel A defines “successful school-to-work transition” more strictly by considering employed individuals without a vocational qualification prior to their employment as unsuccessful. Panel B restricts the sample to individuals on the labor market by excluding those who have completed a further degree after secondary school to rule out that this is driving our results. Finally, Panel C combines both restrictions taken in the approaches shown in panels A and B. Crucially, all point estimates remain relatively close to zero, and nowhere can we reject the null hypothesis of a null effect.

Three patterns emerge from these results. Most importantly, comporment grading neither enhances nor reduces the chances of a successful school-to-work transition. Second, C/S and TWFE estimates hardly differ, irrespective of whether

²¹This approach is visualized in Figure A5.

controls are included. This suggests that treatment effect heterogeneity is not an issue here. This is also vividly illustrated in Figure 3.2, which shows that post-reform point estimates differ neither across time periods nor groups. For this reason, we will adhere to the TWFE approach for the remaining analyses as it allows us to account for the small number of clusters when conducting inference. Furthermore, for the analysis of non-cognitive skills and student achievement, we use eight further federal states to increase statistical power. Ultimately, conclusions from estimated effects are robust to different definitions of the outcome variable and sample restrictions.

Table 3.5: Effect of comporment grading on school-to-work transitions

	Successful School-to-work Transition			
	Callaway & Sant'Anna (C/S)		Two-Way-Fixed-Effects (TWFE)	
	(1)	(2)	(3)	(4)
<i>Panel A: Main</i>				
	0.0028	-0.0022	-0.0055	-0.0062
	[-0.0413, 0.0469]	[-0.0472, 0.0428]	[-0.0400, 0.0290]	[-0.0370, 0.0246]
WCB p-val.	-	-	0.6446	0.6236
<i>Panel B: Four period setup</i>				
	-0.0020	-0.0033	-0.0062	-0.0069
	[-0.0331, 0.0291]	[-0.0327, 0.0262]	[-0.0402, 0.0278]	[-0.0372, 0.0234]
WCB p-val.	-	-	0.6326	0.6116
Mean Dep. Var.	0.88	0.88	0.88	0.88
N (A – B)	16,982	16,982	16,982	16,982
Controls	No	Yes	No	Yes
Std. Error	Robust	Robust	Cluster	Cluster

Notes: Estimates of the overall ATT (see equation 3.3 and 3.4) according to Callaway and Sant'Anna (2021) (columns 1 and 2) and from TWFE regressions using state and cohort fixed effects (columns 3 and 4). Columns 2 and 4 additionally include a female and migration background indicator as control variables. Columns 1 and 2 report simultaneous 95% confidence intervals robust to heteroskedasticity. Columns 3 and 4 report 95% confidence intervals based on cluster-robust standard errors and *p*-values from the wild cluster bootstrap routine using weights from Webb's distribution (Roodman et al. 2019) and 999 iterations.

Source: Mikrozensus waves 2011–2018.

3.5.2 Non-Cognitive Skills and Reading Skills

Having established that grading social and work behavior does not alter school-to-work-transitions, we analyze whether non-cognitive skills and reading skills as intermediate outcomes are affected. Columns 1 to 3 of Table 3.7 show the results of estimating equation 3.1 using z-scored measures of non-cognitive skills as

Table 3.6: Effect of comporment grading on school-to-work transitions using 12 states

	Successful School-to-work Transition			
	Callaway & Sant'Anna (C/S)		Two-Way-Fixed-Effects (TWFE)	
	(1)	(2)	(3)	(4)
	0.0028	-0.0022	0.0273	0.0277
	[-0.0413, 0.0469]	[-0.0472, 0.0428]	[-0.0045, 0.0591]	[-0.0042, 0.0597]
WCB p-val.	-	-	0.2032	0.1712
N	16,982	16,982	43,851	43,851
SE	Robust	Robust	Cluster	Cluster
Controls	No	Yes	No	Yes

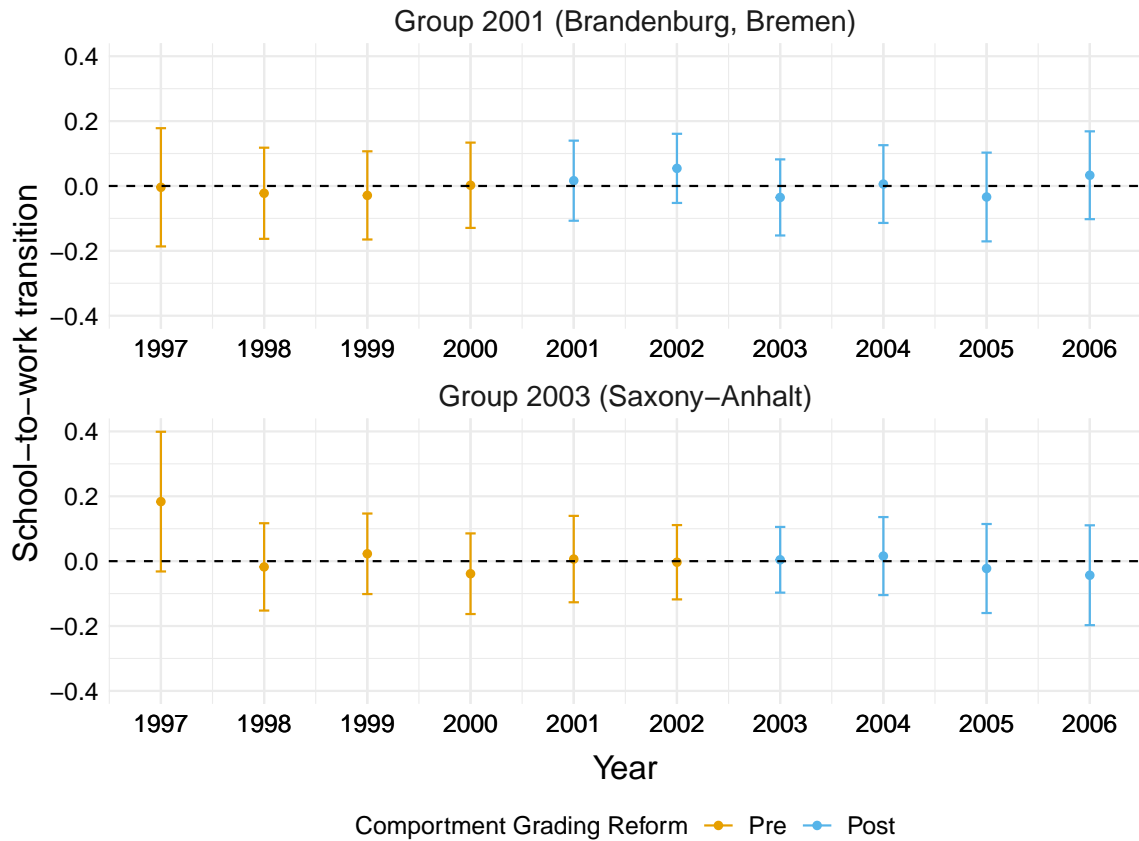
Notes: Estimates of the overall ATT (see equation 3.3 and 3.4) according to Callaway and Sant'Anna (2021) (columns 1 and 2) and from TWFE regressions using state and cohort fixed effects (columns 3 and 4). Columns 1 and 2 use 4 federal states while Columns 3 and 4 use 12 states. Columns 2 and 4 additionally include a female and migration background indicator as control variables. Columns 1 and 2 report simultaneous 95% confidence intervals robust to heteroskedasticity. Columns 3 and 4 report 95% confidence intervals based on cluster-robust standard errors and p -values from the wild cluster bootstrap routine using weights from Webb's distribution (Roodman et al. 2019) and 999 iterations.

Source: Mikrozensus waves 2011–2018.

outcomes from our SOEP sample. In line with the null-effect finding on the school-to-work transition, we do not detect statistically significant effects of comporment grading on any of the non-cognitive skill measures. p -values from the wild cluster bootstrap procedure bolster this finding.

The point estimates displayed in Table 3.7 range from one to below three percent of a standard deviation in absolute value.

Next, we test an alternative potential mediator for long-term effects of comporment grading and investigate whether comporment grading affects reading skills by the end of ninth grade. Column 4 of Table 3.7 reports estimates from equation 3.1 using OLS. Reading test scores are z -scored and have five different plausible values available. We use the procedure from Macdonald (2008) that relies on Rubin (1987). Effectively, it combines estimates from five separate regressions with the respective plausible values and also considers the imputation error emerging from the stochastic nature of plausible values. In line with results from our other outcomes, estimated effects on reading test scores are statistically indistinguishable from zero. Our confidence intervals allow us to reject positive effect sizes on reading test scores beyond roughly .25 SDs, which fall in the category of large effect sizes according to the classification of Kraft 2020. Using the rule of thumb

Figure 3.2: Dynamic effect of compartment grading on school-to-work transitions

Notes: Figure displays estimates of period- and group-specific ATTs for the two treatment groups. The dependent variable is binary and indicates a successful school-to-work transition (see Section 4.3). Specifications include indicators for students' sex and migration background. Error bars correspond to simultaneous 95% confidence bands based on robust standard errors.

Sources: Mikrozensus waves 2011–2018

presented in Woessmann (2016) that average student learning in a year is equal to about a .3 SDs test score increase, we can reject effect sizes that can be translated into learning of 83% of a school year or more.

Tables A7 and A8 in the appendix contain results of robustness checks of our estimates on non-cognitive skills and reading skills. They show that these patterns are robust against a variety of concerns. While Table A7 displays results of regressions without any control variables, Table A8 changes the assignment of treatment based on the year of secondary school enrollment as in Panel D of Table A6. Results in all three cases corroborate our null-effect finding.

Table 3.7: Effect of comporment grading on non-cognitive skills and reading skills

	Trust	Conscientiousness	Agreeableness	Reading Skills
ATT	0.0151 [-0.0808, 0.1110]	-0.0241 [-0.1078, 0.0596]	0.0269 [-0.0541, 0.1078]	-0.0159 [-0.2779, 0.2461]
WCB p-val.	0.7746	0.5737	0.5149	-
Observations	5547	5547	5547	128,249
Adj.R.squared	0.0204	0.0421	0.0118	0.188
Std.Error	Cluster	Cluster	Cluster	Cluster

Notes: Each column presents separate OLS coefficient estimates with federal state and cohort fixed effects. All outcomes are standardized to have mean zero and unit standard deviation. Controls include student sex and a dummy for migration background in columns 1 to 3. In column 4, controls include student sex, migration background, age in months, and an indicator for low SES. Column 4 reports the average estimate across five regressions using separate plausible values of individual reading test scores as implemented by Macdonald 2008. Robust standard errors allow for clustering at the federal state level; wild cluster bootstrap *p*-values and confidence intervals use weights from Webb’s distribution and rely on 9999 iterations in columns 1 to 3 (Roodman et al. 2019). 95% confidence intervals are in box brackets.

Sources: SOEP-Core v37; PISA 2000, 2003, and 2006, IQB-LV 2008-9 (v2), PISA 2012, IQB-BT 2015 (v5)

3.5.3 Discussion of Main Results

Given that only 3 of 16 German federal states introduce comporment grading during our sample period, the available variation for identifying causal effects of this policy is limited. This leads to a rather large degree of statistical uncertainty in our estimates. We are nonetheless able to reject reasonably-sized effects for two of our three main sets of outcomes.

To fix ideas on what we consider to be small and large effect sizes, we rely on the classification put forward in Kraft 2020 and Kraft 2023). In framing benchmarks for effect sizes in the context of evaluating the impact of pre-K–12 education interventions on student achievement, Kraft 2020 suggests the following thresholds: Small ($< .05$ SD), Medium ($.05$ to less than $.2$ SD), and Large ($> .2$ SD). Kraft 2020 derives these classifications from an examination of the distribution of 1,942 effect sizes obtained from the evaluation of 747 Randomized Controlled Trials (RCTs).

This classification can be applied to put our point estimates and their variance estimates into perspective. More specifically, point estimates suggest that comporment grading does not meaningfully affect students’ school-to-work transition. Non-rejectable effect sizes are larger than one tenth of a standard deviation or 3.3

percentage points in absolute value using our preferred specification (the standard deviation of the variable successful school-to-work transition is 0.34; see Table 1).

For non-cognitive skills and reading skills as potential intermediate outcomes, the situation is more ambiguous. Point estimates are close to zero for all skill measures. However, while confidence intervals for non-cognitive skill measures exclude effect sizes beyond 11 percent of a SD (i.e., medium-sized effects according to Kraft 2020), we cannot reject effect sizes on reading skills that amount to less than 25 percent of a SD or, equivalently, 83% of achievement gains during a typical school year (Woessmann 2016). Put differently, we cannot reject even large effect sizes ($> .2$ SD) in the terminology of Kraft 2020.

3.6 Potential Explanations

Our analysis has established that comportment grades do not meaningfully alter student outcomes, but leaves open the question as to why this is the case. Section 4.2 delineates why the effect of comportment grading on student outcomes is ambiguous in theory. Similarly, practitioners disagree about the usefulness of comportment grades. On the one hand, teachers suggest that comportment grading has at best a small effect on various behavioral dimensions and academic achievement. For instance, Table 3.10 shows that more than 75% of respondents see no effect on students' thirst for knowledge, while respondents are split on the question whether there is a positive or no effect on academic performance. On the other hand, employers' fierce insistence on their usefulness to screen applicants (Tuch 2000), which has also been shown in a correspondence study by Protsch and Solga (2015), suggests that receiving comportment grades could indeed affect student outcomes. This section therefore explores the potential channels through which comportment grades could affect student outcomes using teacher survey data as well as additional analyses using NEPS and IQB data.

Direct signaling effect Even in the absence of an effect on reading and non-cognitive skills, comportment grades could still help students signal non-cognitive abilities in the application process and therefore ease their school-to-work transition (Protsch and Solga 2015). However, finding such an effect rests on the assumption that comportment grades provide additional information to employers beyond other parts of the application. Analyzing variation in comportment grades using

simple linear regressions, we find that once subject grades and overall GPA are accounted for, including measures of the two non-cognitive skills conscientiousness and agreeableness adds little explanatory power. More specifically, Table 3.8 shows that the adjusted R^2 increases by a mere 16% (2.7 percentage points) going from column 1 to 4.²² Put differently, the informational value-added of comportment grades about non-cognitive skills relevant for labor market outcomes is limited once subject grades are known. Therefore, we should not expect a large effect on students' school-to-work transition operating through this signaling mechanism.

Our finding regarding the informational content of subject grades is in line with the literature. In particular, subject grades are found to capture various aspects of personality in addition to IQ that should similarly be captured in comportment grades, too. For example, more conscientious individuals take assignments more seriously, which leads to better grades (Borghans et al. 2016). For the same reason, subject grades are more predictive of labor market outcomes than IQ. Furthermore, Ferman and Fontes (2022, p.1) show that teachers "inflate grades of well-behaved students and deduct points from worse-behaved ones" on high-stakes achievement tests in Brazil, supporting the notion that subject grades contain information about student behavior.

Finally, teachers in our survey are split regarding the informational content of comportment grades: 33% of teachers (strongly) agree that comportment grades are already contained in subject grades, while 52% (strongly) disagree (see Figure 3.3).

Indirect effect mediated through reading and non-cognitive skills A possible effect of comportment grades on the school-to-work transition could also be mediated by reading and non-cognitive skills. However, there are two major theoretical arguments as to why it is reasonable to expect null effects. First, comportment grades as implemented in Germany and other countries provide a low-stake incentive to behave better as they typically do not count towards tracking decisions and the promotion to the next grade. It is therefore expected that students do not exert much effort to obtain better comportment grades (e.g., Schlosser et al. 2019). Yet, student effort is found to be an important input in the education production function (Stinebrickner and Stinebrickner 2004; De Fraja et al. 2010;

²²Note that most of the variance in comportment grades remains unexplained. This could be due to measurement error or because there are many important determinants of comportment grades that are missing in our data.

Table 3.8: Explaining Comportment Grades using Subject Grades

Explanatory Variable	Dep. Var.: Comportment Grades			
	(1)	(2)	(3)	(4)
German Grade	0.124 (0.034)	0.129 (0.034)	0.119 (0.034)	0.123 (0.034)
Math Grade	-0.006 (0.029)	-0.005 (0.029)	-0.008 (0.029)	-0.007 (0.029)
GPA	0.328 (0.046)	0.326 (0.046)	0.324 (0.046)	0.322 (0.046)
Agreeableness		0.052 (0.023)		0.040 (0.023)
Conscientiousness			0.113 (0.022)	0.108 (0.022)
Constant	1.957 (0.373)	1.991 (0.369)	2.081 (0.380)	2.103 (0.377)
Controls	Yes	Yes	Yes	Yes
Observations	886	886	886	886
R2 Adj.	0.167	0.171	0.192	0.194

Notes: Each column presents separate OLS coefficient estimates with robust standard errors in brackets. Controls include student age and gender. Comportment Grade, German Grade, Math Grade, and GPA are rounded to take on integers from 1 (worst) to 6 (best). Measures of Conscientiousness and Agreeableness are z-scored.

Sources: NEPS SC3 11.0.0.

Gneezy et al. 2019; Rury and Carrell 2023). Therefore, student outcomes should not be affected through this channel.

Second, the biannual – or even annual – release of report cards with comportment grades in Germany may not provide feedback that is appropriate to change students' behavior. Previous research by Levitt et al. (2016) shows that students no longer respond to performance incentives once rewards are provided with a delay. Moreover, Jalava et al. (2015) demonstrate that giving numerical or letter grades may not be effective to incentivize students, whereas giving symbolic rewards or providing relative rank information could be more effective. Similarly, teachers might provide feedback regarding students' behavior through other means than comportment grades, e.g., pedagogical disciplinary concepts such as reprimands.

Empirically, we can provide indirect evidence to bolster these arguments by using two separate data sources. As the effect of comportment grades on reading skills might operate through a less disruptive classroom environment (Lazear

2001; Kristoffersen et al. 2015; Ahn and Trogdon 2017), we use cross-sectional data from the first PISA wave 2000 and our own teacher survey to analyse whether comportment grades are associated with better classroom discipline and teachers' disciplinary measures.

Table 3.9: Contemporaneous comportment grading and disciplinary problems as rated by school principal and classroom disruptions as rated by students

	Classroom disruptions	Disciplinary problems in class
Comportment Grading	-0.0307 [-0.0950, 0.0336]	-0.0834 [-0.2199, 0.0531]
Student + classroom cont.	yes	-
Principal cont.	-	yes
School cont.	yes	yes
Outcome mean	-0.01	0.02
R-squared	0.013	0.193
Observations	22,761	1,013

Notes: Each column presents a separate OLS coefficient estimate from a cross-sectional regression of the contemporaneous comportment grading policy in the 16 federal states on the frequency of learning impairments at school, measured as a standardized index of six questions answered by students in column (1) and on the existence of learning impairments at school, measured as a standardized index of six questions answered by school principals in column (2). "Comportment Grading" is binary and equals 1 if state s had comportment grading in place. Student controls include gender, age, migration background, and parental education. Classroom controls include the student body composition regarding the same variables. Principal controls include principal's gender, age, and years of experience as a teacher. School controls include school type, school size, public school dummy, and school location. 95% confidence intervals in box brackets. Heteroskedasticity-robust standard errors used.

Source: PISA 2000.

Relying on data from the first PISA wave 2000, we use the frequency of classroom disruptions as a proxy for how conducive the classroom environment is to foster learning.²³ Column 1 of Table 3.9 uses an index combining the answers of students

²³Please note that the analyzed sample, drawn from PISA data from 2000, comprises four states where students did not receive comportment grades contemporaneously and eight states where students did receive comportment grades, as depicted in Figure 3.1. We refrain from utilizing PISA data from later waves due to the availability of only one state without comportment grades remaining from 2003 onward.

to questions about classroom disruptions in German classes.²⁴ We standardize the outcome to have mean zero and unit standard deviation. The results in Table 3.9 show that there is no correlation between a state's contemporaneous comportment grading policy and the number of classroom disruptions as assessed by students.

Column 2 of Table 3.9 uses PISA 2000 data to analyze the association between comportment grading and disciplinary problems in class reported by school principals. The outcome variable is an index derived from a set of questions²⁵ which we standardize to have mean zero and unit standard deviation. The results in Table 3.9 show that there is no correlation between a states' contemporaneous comportment grading policy and the number of classroom disruptions as assessed by school principals. In summary, the cross-sectional evidence derived from PISA 2000 fails to reveal a significant correlation between comportment grading and classroom behavior. However, it is noteworthy that the correlation has the expected sign: Comportment grading is negatively associated with classroom disruptions and disciplinary problems in class, albeit insignificantly so.

For additional insights into the connection between comportment grading and student behavior, we turn to data obtained from our teacher survey eliciting information on the implementation of disciplinary measures. Table 3.10 tentatively indicates that teachers tend to employ the same number of alternative disciplinary measures, such as one-to-one conversations with disruptive students, when comportment grading is in use versus when it is not.

The interchangeability observed between various disciplinary approaches and comportment grading indicates that the feedback students receive about their behavior remains consistent, irrespective of the presence of comportment grades. Consequently, the introduction of comportment grades is unlikely to yield a positive impact on students' classroom behavior. However, improved classroom behavior could potentially have direct positive effects on both non-cognitive skills and reading skills. Consequently, neither outcome is likely to be impacted by the presence or absence of comportment grades.

²⁴The six questions are as follows. "teacher has to wait until students become quiet", "students cannot work undisturbed", "students do not listen to teacher", "students only start working long after beginning of class", "it is noisy in class", "nothing happens in the first five minutes". Each item can be answered by the students with "never" (1), "in few lessons" (2), "in most lessons" (3), "in all lessons" (4).

²⁵The questions are as follows. "classroom disruptions by students", "student truancy", "students lacking respect for teachers", "bullying of students by classmates", "frequent absence of student", "lacking parental support when learning at home". Each item can be answered by the school principal with "not impaired" (1), "little impaired" (2), "somewhat impaired" (3), "very impaired" (4).

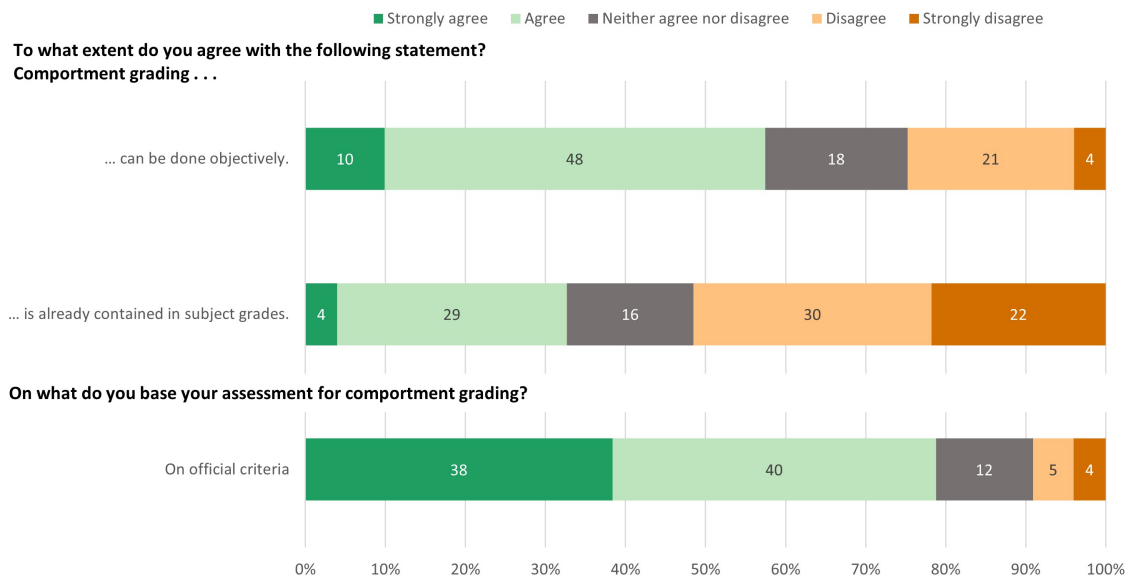
Table 3.10: Teacher Survey

Aspect of Comportment Grading	Share
Panel A: Effect on (student) outcomes	
Thirst for knowledge	
Positive effect / No effect / Negative effect	20% / 77% / 3%
Academic performance	
Positive effect / No effect / Negative effect	51% / 47% / 2%
Discipline in class	
Positive effect / No effect / Negative effect	63% / 36% / 1%
Panel B: How many teachers are involved?	
Classroom teacher only	8%
Classroom teacher and others	92%
<i>Conditional on more than one teacher</i>	
1–5 / 6–10 / ≥ 11	36% / 48% / 16%
Panel C: Minutes taken per pupil and report card	
0–9 / 10–19 / ≥ 20	28% / 31% / 41%
Panel D: Assessment considered reasonable	
Reliability & diligence / Willingness to learn & perform	80% / 78%
Cooperative & team spirit / Conflict skills & tolerance	76% / 71%
Panel E: Need other measures more in absence of	
Yes	58%
No	42%
<i>Conditional on Yes</i>	
1-to-1 conv. / Warning / Class register entry	36% / 26% / 20%

Notes: Based on 245 responses. Panel A is based on the question “To what extent do comportment grades in report cards have an effect on ...?”. Panel D is based on the question “In which behavioral dimensions do you consider an assessment reasonable?”. Panel E is based on the question “Please imagine there were no comportment grades. Would you have to make use of the following measures to sanction inadequate behavior more often?”. Panel D and Panel E allowed for multiple answers.

Source: Own survey among German teachers.

Summary and limitations In sum, both theoretical and empirical arguments suggest that our null findings are meaningful. We provide evidence that comportment grades on report cards do not provide significant additional information relevant for labor market outcomes, making a direct signalling effect unlikely. In addition, we show that comportment grades are not associated with a better learning environment in the classroom, which is evidence against a direct effect

Figure 3.3: Teacher survey: Assessment and basis of comportment grading

Notes: Based on 245 responses.

Source: Own survey among German teachers.

on reading skills. Finally, teachers report that they provide students with similar feedback on their behavior in the classroom irrespective of whether a comportment grading policy is in place, implying that students probably do not change their behaviors in different ways under the two scenarios.

Although we lack the data and variation to empirically test the theoretical arguments surrounding the low-stakes nature of the incentives provided by comportment grades and the importance of timely feedback in behavior modification, both of them suggest that comportment grades, as implemented in German schools, are unlikely to affect student outcomes.

3.7 Costs of Comportment Grading

The costs of the reform primarily consist of the time invested by teachers. To present a back-of-the-envelope estimate of these costs, we use our teacher survey to elicit information about the time investments teachers make to determine comportment grades. Regarding the number of teachers involved in determining comportment grades, 92% of the teachers in our survey report that more than one teacher is involved in comportment grading (see Table 3.10). When asked to provide exact

numbers, most respondents state that for each class, there are 6 to 10 other teachers involved in grading in addition to the class teacher. This range is broadly in line with the number of school subjects included in a timetable for a class. The range therefore indicates that all teachers who give lessons to the class in any subject are typically part of the process. To arrive at a conservative estimate, we only focus on the time costs born by the class teacher, who usually does the main work.

How much time do teachers spend on the grading? Our expert survey uncover considerable heterogeneity in the time invested in comporment grading. Most respondents take fewer than 30 minutes per student and report card. For our cost calculations, we take the midpoint of the median answer, which is 15 minutes per student and report card (see Table 3.10). Given that comporment grading is usually conducted for both half-term and end-of-year certificates, they invest about 30 minutes each year. This amount of time seems high at first. However, several survey participants point out that the documentation of students' behavior occurs throughout the entire school year. That is, the preparatory work to determine the final grade starts well before the actual writing of the report card. Moreover, the work also includes the time to coordinate the grading with other teachers who give lessons to the same students.

The final step is to valuate the 30 minutes spent on comporment grading per student and year. According to the OECD Education Statistics, a teacher's average salary in Germany is about \$88,071 (gross) per year. Assuming a teacher typically works 40 hours per week, we arrive at an estimated cost of \$21.3 per student and year. For the roughly 11 million students in Germany, this adds up to a total annual cost of about \$235 million. This amount is sufficient to, for instance, finance virtual coaching programs for students (Oreopoulos et al. 2020) or to run information campaigns aimed at improving student behavior (see Peter et al. 2021, for a related campaign in the German context).

3.8 Conclusion

Exploiting limited policy variation across German federal states, we document that grading students' comporment in school does not meaningfully affect their success in transitioning from school to work. We can reject that receiving comporment grades implies an increase or decrease of more than 3.3 percentage points or, equivalently, 10 percent of a SD, in the probability of successfully transitioning

into the labor market. In line with this finding, point estimates for non-cognitive and reading skills are close to zero, with the caveat that the large amount of statistical uncertainty in the case of reading skills does not allow us to exclude even substantial effect sizes.

For benchmarking purposes, we use the classification proposed by (Kraft 2020; Kraft 2023) Notably, we can exclude medium-sized effects (i.e., <0.2 SD) on students' school-to-work transition and non-cognitive skills. However, for reading skills, we cannot rule out large effect sizes (i.e., >0.2 SD).

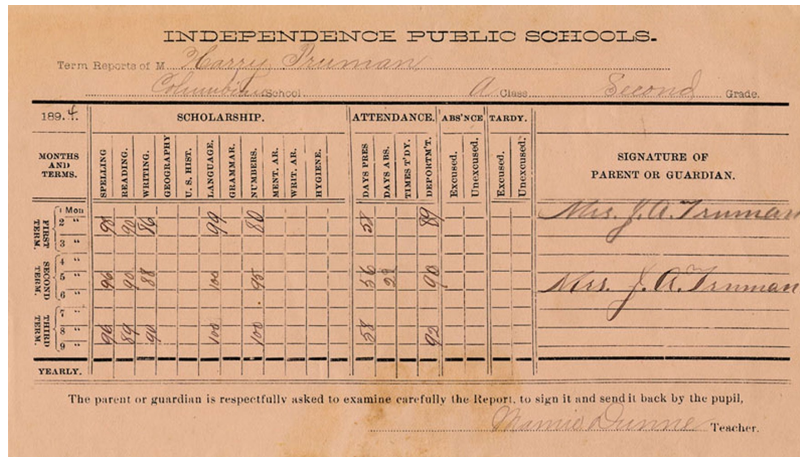
Our robustness checks further bolster our findings: Using alternative estimation strategies, including different sets of control variables, and applying other sample restrictions hardly affects our results. Finally, we explore potential explanations for the null result and present both theoretical and empirical evidence that compartment grades lack the effectiveness needed to significantly affect student outcomes. To arrive at this conclusion, we adopt a careful approach to identification and estimation and rely on numerous data sources.

Our findings suggest that the arguments of neither proponents nor opponents of compartment grading can be supported by causal evidence. We also provide supplementary evidence on why this might be the case. However, there are two caveats to this study. First, limited reform variation implies elevated statistical uncertainty, which is why we distinguish among our results to which extent they allow to rule out potential effect sizes. Second, whether compartment grading is an input to the education production function might differ across countries. Yet, given that compartment grading policies are similar in other countries (see Table A1), we remain confident in the external validity of our results, meaning that they might be informative for other countries that consider the introduction or abolition of compartment grading.

In sum, this paper shows that the introduction of compartment grading does not have an effect on student outcomes, with the caveat that we cannot rule out even large effect sizes for reading skills. Finding null effects of educational reforms is not uncommon (e.g. Dale and Krueger 2002; Fryer 2011; Jerrim et al. 2018; Leuven and Løkken 2020; Bird et al. 2021). At the same time, this is highly informative from a policy perspective: It is crucial to know whether much-debated reforms affect student outcomes at all, especially when they incur significant costs. Our null-effect finding on students' school-to-work transition and their non-cognitive

skill formation suggests that policy efforts should focus on other domains to increase the efficiency of the education system.

Figure A2: School report card for Harry S. Truman



Notes: Truman (1884–1972): second-grade report card, includes grade for “department” (synonym for “comportment”, last item within the “attendance” category).
 Source: Harry Truman Library.

Figure A3: Stylized overview of the German school system

Years of schooling				Age	
			University entrance	19	Secondary school
13	Vocational training/ further schooling/ labor market	Vocational training/ technical college/ labor market	Academic track (Gymnasium)	18	
12				17	
11	16				
10	15				
9	14				
8	13				
7	12				
6	11				
5	10				
4	Primary school (Grundschule)			9	Primary school
3				8	
2				7	
1				6	

Source: Own representation based on Helbig and Nikolai (2015).

Table A1: Grading of social and work behavior in selected European countries

Country	Grading of work and social behavior
Austria	Behavioral grades exist for all school types and grade behavior in the middle school years. In 2014, parents' associations tried to abolish these grades (<i>Die Presse</i> , Sept. 18, 2014).
Czech Republic	Students' behavior is assessed as (1) very good, (2) satisfactory, or (3) unsatisfactory.
Denmark	Until 2013, students received grades on the orderliness/organization/neatness of their written exams in Danish and mathematics (Landersø and Heckman 2017).
France	A grade for comportment ("note de vie scolaire") was abolished in 2014. The grade considered punctuality, respect for rules, participation in the school's social life, and attaining a road safety education certificate. It was abolished following criticism regarding its subjectivity (<i>Avis du Conseil supérieur des programmes sur la note de vie scolaire</i> , Nov. 21, 2013).
Greece	At the end of each quarter and when grades have been finalized and recorded, parents receive an individual progress report and are informed about student performance, diligence, attendance and behavior.
Hungary	Behavior and effort/diligence are evaluated on a four-grade scale: exemplary (5), good (4), varying (3), or poor (2).
Italy	The assessment of students' conduct refers to the development of citizenship competences, in accordance with what is established by each school's regulations and the 'Joint responsibility agreement' signed by students and parents. Students with a mark below 6/10 in conduct cannot progress to the following grade.
Norway	The students are assessed in conduct.
Poland	A grade for behavior exists and does not influence the promotion to a higher grade or graduation. Yet, receiving an inadmissible grade for behavior in two consecutive years student cannot be promoted to the next grade or finish school.
Sweden	A proposal to reintroduce comportment grading in schools caused a long debate in 2019. A majority of members of the Riskdag upheld the proposal with the aim of reducing disruptive behavior in schools. The Swedish Teachers' Association is critical and fears that grading conduct might even be counterproductive (<i>Göteborgs-Posten</i> , Apr. 2, 2019).
Switzerland	Social conduct and attitude to work may be assessed depending on canton. In 2016, the canton of Zurich also decreed that these grades count towards students' promotions to high-track schools (<i>Tages-Anzeiger</i> , Dec. 19, 2016).

Source: European Commission (2021). *Eurydice: Better knowledge for better education policies*. National Education Systems. Individual country reports retrieved from https://eacea.ec.europa.eu/national-policies/eurydice/national-description_en (as of July 5, 2021).

Table A2: Teacher guidelines for the evaluation of behavior (excerpt) in the state of Baden-Wuerttemberg

Criterion	Commendable behavior	Gross misconduct
General conduct	Polite, friendly, controlled, calm, placid	Naughty, defiant, malicious, uncontrolled, quick-tempered
Camaraderie	Companionable, helpful, compassionate, compatible	Non-companionable, ruthless, unbearable, spiteful
Honesty	Sincere, honest, candid	Insincere, dishonest, lying
Restraint	Modest, restrained, discreet	Immodest, boastful, presumptuous, arrogant
Work effort	Takes over community tasks willingly	Refuses to take over community tasks
Acceptance of rules	Recognition of principles of order, sense of order, willingness to comply, reliable, punctual, regular participation in class, compliant	Negligently or intentionally violates principles of order, disorganized, belligerent, unreliable, frequently arrives late, frequently misses class without sufficient justification, continually disrupts class

Notes: This table was suggested as a teacher aide to assess student behavior in the state of Baden-Wuerttemberg. Most commonly, students receive the grade “good”. If the student’s behavior is particularly cooperative, the grade “very good” might be assigned. If the student’s behavior frequently meets the description given by the columns *Misconduct* (not shown in this excerpt) or *Gross misconduct*, the student might receive a “satisfactory” or “insufficient” grade. *Source:* Hausmann, Johanna (2010). *Beeinflussungstendenzen bei Kopfnoten: welche Faktoren fließen in die Noten unserer Kinder ein?* Hamburg, Diplomica.

Table A3: Teacher guidelines for the evaluation of behavior in the state of Saxony

Criterion	Behaviors to be considered
Order	Care, punctuality, reliability, compliance with rules, having teaching materials ready
Cooperation	Initiative, willingness to cooperate, ability to work in a team, independence, creativity, responsibility
Conduct	Attentiveness, helpfulness, civic courage and appropriate handling of conflicts, considerateness, tolerance, sociability, self-perception
Diligence	Willingness to learn, determination, endurance, regularity in fulfilling task.

Notes: This table represents the concept of comportment grading in the state of Saxony. Students will be assigned a grade between 1 (“exemplary”) and 5 (“insufficient”). *Source:* Bohl, Thorsten (2010). “Aktuelle Regelungen zur Leistungsbeurteilung und zu Zeugnissen an deutschen Sekundarschulen”. In: *Zeitschrift für Pädagogik* 49.4, p. 558.

3.B Appendix B: Treatment Effect Estimands

Following the exposition by Callaway and Sant'Anna (2021), this section details how the ATTs we report in columns 1 and 2 of Table 3.5 are obtained. Let G_i be the time period when unit i becomes treated and $t = 1, \dots, T$ denote time periods. $Y_{it}(g)$ is unit i 's potential outcome in time period t if they become treated in period g .

Under (conditional) parallel trends and for all $t \geq g$, they show that the following group- and period-specific average treatment effect on the treated is identified using modified differences in expectations

$$\text{ATT}(g, t) := E(Y_t(g) - Y_t(0) | G = g).$$

Effects with $t < g$ can be used for pre-testing. In the canonical 2x2 design, $\text{ATT}(g = 2, t = 2)$ is the estimand of interest, corresponding to the instantaneous treatment effect for the group receiving treatment in the second period. In general staggered designs with many more ATTs, aggregates of these can be used to get an idea of the overall treatment effect.

In our setup, units correspond to German federal states, i.e. $i \in \{\text{Brandenburg, Bremen, Saxony-Anhalt, North Rhine-Westphalia}\}$. We restrict the sample period to $t = 1992, \dots, 2009$. Note that we lack a control group in 2007 and later. There are two treatment groups receiving treatment in 2001 and 2003, respectively ($g \in \{2001, 2003\}$). This means that we have 14 ATTs for each group, 8 (10) pretreatment and 6 (4) post-treatment effects for the group with $g = 2001$ ($g = 2003$). In a first step, we average over post-treatment effects for each group:

$$\begin{aligned} \text{ATT}_S(g = 2001) &:= \frac{1}{2006 - 2001 + 1} \sum_{t=1993}^{2009} 1_{\{2001 \leq t \leq 2006\}} \text{ATT}(g = 2001, t) \\ &= \frac{1}{6} \sum_{t=2001}^{2006} \text{ATT}(g = 2001, t) \end{aligned}$$

$$\text{ATT}_S(g = 2003) := \frac{1}{4} \sum_{t=2003}^{2006} \text{ATT}(g = 2003, t).$$

To arrive at a single measure that resembles a multi-group multi-period extension

of the ATT in the 2x2 design, we further average across treatment groups to obtain

$$\begin{aligned} \text{ATT} &:= \sum_{g \in \{2001, 2003\}} \text{ATT}_S(g) \cdot \Pr(G = g) \\ &= \text{ATT}_S(g = 2001) \cdot \Pr(G = 2001) + \text{ATT}_S(g = 2003) \cdot \Pr(G = 2003). \end{aligned} \quad (3.3)$$

Table 3.5 shows estimates of the estimand in equation 3.3 under different scenarios.

As outlined in the empirical strategy (Section 3.4), the simultaneous confidence bands account for the estimation uncertainty along the entire path of group-time effects. This implies that confidence intervals widen in the number of estimated group-time effects. To maximize power when estimating the ATT in equation 3.3, one can reduce the number of estimated effects in the following way without loss of generality. The reforms and our sample periods give rise to four time periods: before the first group gets treated (1992 to 2000), after the first and before the second group gets treated (2000 to 2002), after the second group got treated and before North Rhine-Westphalia adopts the treatment (2003 to 2006), and 2007 and later. They are denoted by $t^* = 1, \dots, 4$. Note that we lack a control group in period 4. Following the same logic as above, this gives 0 (1) pre-treatment and 2 (1) post-treatment effects for the group with $g = 2001$ ($g = 2003$). Although this setup is more akin to classical difference-in-differences (e.g., Card and Krueger 1994), having a second treatment group allows us to use one period to pre-test for parallel trends. The averaging of effects then takes the following form:

$$\begin{aligned} \text{ATT}_S(g = 2) &:= \frac{1}{2} \sum_{t^*=1}^4 1_{\{2 \leq t^* \leq 3\}} \text{ATT}(g = 2, t^*) \\ &= \frac{1}{2} \sum_{t^*=2}^3 \text{ATT}(g = 2, t^*) \end{aligned}$$

$$\text{ATT}_S(g = 3) := \text{ATT}(g = 3, t^* = 3).$$

To arrive at a single measure that resembles a multi-group multi-period extension of the ATT in the 2x2 design, we further average across treatment groups to obtain

$$\begin{aligned} \text{ATT} &:= \sum_{g \in \{2, 3\}} \text{ATT}_S(g) \cdot \Pr(G = g) \\ &= \text{ATT}_S(g = 2) \cdot \Pr(G = 2) + \text{ATT}_S(g = 3) \cdot \Pr(G = 3). \end{aligned} \quad (3.4)$$

3.C Appendix C: Teacher Survey

Table A4: Descriptive statistics teacher survey: Teacher characteristics

	Sample percentage	National percentage
School type		
Primary school	25.31	31.2
Basic school (Hauptschule)	6.12	4.0
Middle school (Realschule)	23.67	8.1
Academic track (Gymnasium)	22.45	26.4
Integrated compreh. school	14.69	13.8
Others (e.g., vocational schools)	7.76	NA
Place of Work		
West Germany	61.22	73.0
East Germany	38.37	25.3
No answer	0.41	1.7
Age		
29 years or younger	3.27	7.0
30-39 years	26.53	29.0
40-49 years	22.45	26.0
50-59 years	36.73	26.0
60 years or older	10.61	11.0
No answer	0.41	0.0
Gender		
Female	68.57	73.4
Male	29.39	26.6
Diverse	0.41	NA
No answer	1.63	NA
Observations	246	799314

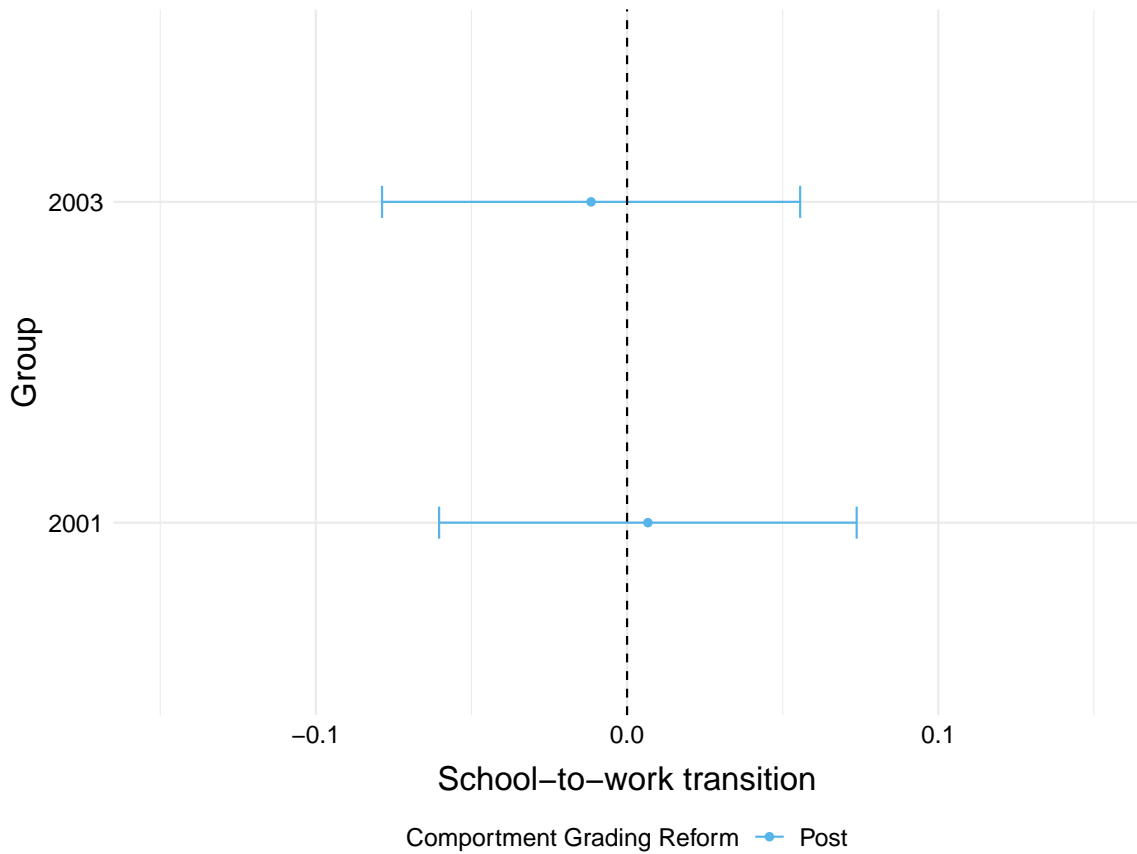
Notes: Sample includes teachers that participated in our online survey. NA indicates that a certain characteristic was not available from official statistics.

Sources: The national percentage is taken from the German Statistical Office's 2021/2022 school report.

3.D Appendix D: Robustness Checks

3.D.1 School-to-Work Transition

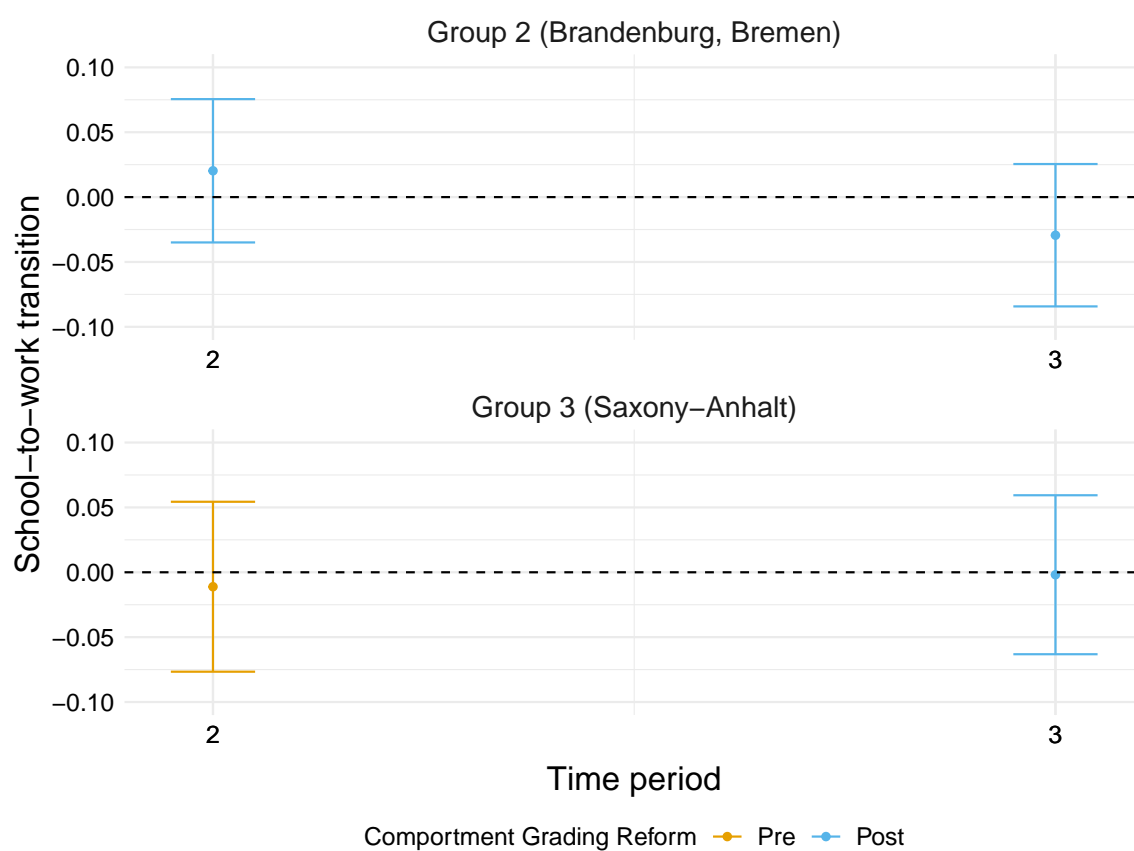
Figure A4: By group treatment effect of comporment grading on school-to-work transitions



Notes: Figure displays estimates of treatment-group-specific effects. The dependent variable is binary and indicates a successful school-to-work transition (see Section 4.3). Specifications include indicators for students’ sex and migration background. Error bars correspond to simultaneous 95% confidence bands based on robust standard errors.

Sources: Mikrozensus waves 2011–2018

Figure A5: Four period setup: Dynamic effect of compartment grading on school-to-work transitions



Notes: Figure displays estimates of treatment-group-specific effects. The dependent variable is binary and indicates a successful school-to-work transition (see Section 4.3). Specifications include indicators for students' sex and migration background. Error bars correspond to simultaneous 95% confidence bands based on robust standard errors.

Sources: Mikrozensus waves 2011–2018

Table A5: Effect of compartment grading on school-to-work transitions: Robustness checks

	Successful School-to-work Transition			
	Callaway & Sant'Anna (C/S)		Two-Way-Fixed-Effects (TWFE)	
	(1)	(2)	(3)	(4)
<i>Panel A: Stricter definition of success</i>				
	−0.0017	−0.0036	−0.0160	−0.0169
	[−0.0500, 0.0467]	[−0.0563, 0.0491]	[−0.0459, 0.0138]	[−0.0427, 0.0088]
WCB p-val.	-	-	0.2763	0.1782
<i>Panel B: Without those catching up</i>				
	0.0242	0.0225	0.0107	0.0081
	[−0.0270, 0.0754]	[−0.0285, 0.0735]	[−0.0251, 0.0464]	[−0.0220, 0.0382]
WCB p-val.	-	-	0.5746	0.6446
<i>Panel C: Panel A and B combined</i>				
	0.0254	0.0305	0.0041	0.0009
	[−0.0298, 0.0806]	[−0.0250, 0.0860]	[−0.0289, 0.0372]	[−0.0251, 0.0269]
WCB p-val.	-	-	0.7137	0.9329
Mean Dep. Var.	0.88	0.88	0.88	0.88
N (A)	16,982	16,982	16,982	16,982
N (B – C)	12,867	12,867	12,867	12,867
Controls	No	Yes	No	Yes
Std. Error	Robust	Robust	Cluster	Cluster

Notes: Estimates of the overall ATT (see equation 3.3) according to Callaway and Sant'Anna (2021) (columns 1 and 2) and from TWFE regressions using state and cohort fixed effects (columns 3 and 4). Columns 2 and 4 additionally include a female and migration background indicator as control variables. Columns 1 and 2 report simultaneous 95% confidence intervals robust to heteroskedasticity. Columns 3 and 4 report 95% confidence intervals based on cluster-robust standard errors and *p*-values from the wild cluster bootstrap routine using weights from Webb's distribution (Roodman et al. 2019) and 999 iterations.

Source: Mikrozensus waves 2011–2018.

Table A6: Effect of comporment grading on school-to-work transitions: Robustness checks using treatment assignment in different grades

Two-Way-Fixed-Effects (TWFE)		
	(1)	(2)
<i>Panel A: First grade treatment assignment (Main)</i>		
	−0.0055	−0.0062
	[−0.0400, 0.0290]	[−0.0370, 0.0246]
WCB p-val.	0.6446	0.6236
N	16,982	16,982
<i>Panel B: Second grade treatment assignment</i>		
	−0.0144	−0.0165
	[−0.0453, 0.0166]	[−0.0450, 0.0120]
WCB p-val.	0.3754	0.3113
N	17,385	17,385
<i>Panel C: Third grade treatment assignment</i>		
	−0.0432	−0.0454
	[−0.0834, −0.0030]	[−0.0852, −0.0056]
WCB p-val.	0.0591	0.0611
N	17,477	17,477
<i>Panel D: Fourth grade treatment assignment</i>		
	−0.0374	−0.0388
	[−0.0848, 0.0099]	[−0.0833, 0.0056]
WCB p-val.	0.0571	0.0631
N	17,486	17,486
SE	Cluster	Cluster
Controls	No	Yes

Notes: Estimates of the overall ATT (see equation 3.3) from TWFE regressions using state and cohort fixed effects. Column 2 additionally includes a female and migration background indicator as control variables. Columns 1 and 2 report 95% confidence intervals based on cluster-robust standard errors and *p*-values from the wild cluster bootstrap routine using weights from Webb's distribution (Roodman et al. 2019) and 999 iterations.

Source: Mikrozensus waves 2011–2018.

3.D.2 Non-cognitive Skills and Reading Skills

Without controls

Table A7: Effect of comportsment grading on non-cognitive skills and reading skills - without controls

	Trust	Conscientiousness	Agreeableness	Reading Skills
ATT	-0.0011 [-0.1071, 0.1048]	-0.0132 [-0.1017, 0.0753]	0.0272 [-0.0586, 0.1130]	0.0033 [-0.2909, 0.2975]
WCB p-val.	0.9846	0.7762	0.5525	-
Observations	5547	5547	5547	128,249
Adj.R.squared	0.0045	0.0085	0.0035	0.023
Std.Error	Cluster	Cluster	Cluster	Cluster

Notes: Each column presents separate OLS coefficient estimates with federal state and cohort fixed effects. All outcomes are standardized to have mean zero and unit standard deviation. Column 4 reports the average estimator across five regressions using separate plausible values of individual reading test scores as implemented by Macdonald (2008). Specifications do not include further covariates. Robust standard errors allow for clustering at the federal state level; wild cluster bootstrap p-values and confidence intervals use weights from Webb's distribution and rely on 9999 iterations (Roodman et al. 2019). 95% confidence intervals in box brackets.

Sources: SOEP-Core v37, PISA 2000, 2003, and 2006, IQB-LV 2008-9 (v2), PISA 2012, IQB-BT 2015 (v5).

Treatment assignment in grade 4

Table A8: Effect of comportment grading on non-cognitive skills and reading skills - Treatment assignment in grade 4

	Trust	Conscientiousness	Agreeableness	Reading Skills
ATT	0.0113 [-0.0717, 0.0944]	-0.0163 [-0.1394, 0.1069]	0.0175 [-0.1067, 0.1418]	0.0414 [-0.1162, 0.1989]
WCB p-val.	0.7957	0.8171	0.7788	-
Observations	5547	5547	5547	128,249
Adj.R.squared	0.0204	0.0421	0.0118	0.046
Std.Error	Cluster	Cluster	Cluster	Cluster

Notes: This specification checks the robustness of the effects assuming that comportment grading was introduced in the fourth rather than the first year. Each column presents separate OLS coefficient estimates with federal state, cohort, and survey year fixed effects. All outcomes are standardized to have mean zero and unit standard deviation. Column 4 reports the average estimator across five regressions using separate plausible values of individual reading test scores as implemented by Macdonald (2008). Controls include student sex and a dummy for migration background in columns 1-3 and in addition age in months and an indicator for parental SES in column 4. Robust standard errors allow for clustering at the federal state level; wild cluster bootstrap p-values use weights from Webb's distribution and rely on 999 iterations (Roodman et al. 2019). 95% confidence intervals in box brackets.

Sources: SOEP-Core v36, PISA 2000, 2003, and 2006, IQB-LV 2008-9 (v2), PISA 2012, IQB-BT 2015 (v5).

3.D.3 Effect on Composition of the Academic Track

Table A9: Effect of compartment grading on academic track school attendance

	(2)
	Academic Track School Attendance
ATT	-0.0783 [-0.2028, 0.0462]
Outcome mean	0.33
Adj. R-squared	0.126
Observations	128,249
St. Error	Cluster

Notes: OLS coefficient estimate with federal state, cohort, and survey year fixed effects. Controls include student sex, migration background, age in months, and an indicator for parental SES. Academic Track School Attendance is an indicator variable. Robust standard errors allow for clustering at the federal state level. 95% confidence intervals are in box brackets.

Sources: PISA 2000, PISA 2003, PISA 2006, IQB-LV 2008-9 (v2), PISA 2012, IQB-BT 2015 (v5).

3.E Appendix E: Federal State of Schooling

To alleviate concerns regarding measurement error in treatment assignment, we restricted our sample to minimize the number of wrongly assigned individuals. By doing so, we only misclassify an average of 2.03% of individuals in the German Socio-Economic Panel (SOEP) sample and of 4.19% in the Mikrozensus sample.²⁶ In the SOEP, we decided to mainly include individuals who participated in the youth surveys. These target adolescents aged 16 to 17 in the survey households, treating them as first-time adult respondents. We add further individuals from the adult surveys aged between 15 and 20. Similarly, the Mikrozensus sample is now restricted to young adults aged between 15 and 20.

For the SOEP, we have three different routes to imputing the federal state in which the individuals enrolled in primary school at our disposal. The first is to simply use the information on respondents' federal state of enrolment or, alternatively, on their state of transition into secondary school, which is available for a subset of respondents. This first strategy allows us to obtain the federal state at enrolment in primary school for about 40% of the sample.

For the remainder, for whom this is not possible, our "second-best" strategy exploits the fact that the households in which the youth respondents live participate in the SOEP themselves. This allows us to retrieve the federal state of respondents' households in those survey years in which respondents were between 6 and 10 years old, i.e., attended primary school. The federal state of enrolment for another 10% of our sample is thus added.

Lastly, for the latter half of our sample, whose federal state of enrolment we did not grasp by any of the previous means, we use the federal state given in the youth questionnaire, that is, at age 17. Reassuringly, the misclassification rate remains low at roughly 3.2% when we compare the federal state at age 17 with that during primary school as obtained in the previous two routes. This is little surprising given that about 98% of 17-year-olds in Germany still live with their parents according to Mikrozensus data for 2019 (see here).

This third route is also how we proceed in the Mikrozensus sample, that is, we use the federal state at the time the survey was conducted. Comparing the federal state at ages 15 to 20, i.e., the age range of our Mikrozensus sample, to that during

²⁶Average misclassification rates are computed separately for each age and then weighted by the age distribution in each sample.

primary school in the SOEP sample produces a misclassification rate of 5.9% at maximum for those aged 20 and an average rate of 4.19% for the entire sample.

4

Worker Beliefs About Firm Training¹

4.1 Introduction

Technological change constantly alters the demand for the skills needed to perform a broad range of job tasks on the labor market, requiring workers to acquire new skills throughout their careers (Autor et al. 2003; Deming and Noray 2020; Autor et al. 2024). Classic models of human capital depict the accumulation of these skills through both formal schooling and on-the-job learning, much of the latter taking place within firms (Becker 1964; Mincer 1974; Acemoglu and Pischke 1998). The empirical importance of firm training, measured in terms of its contribution to the total human capital stock, is substantial and at least rivals that of formal schooling (Black et al. 2023). However, firm training participation varies widely across workers, and the underlying determinants of firm training decisions at the worker level are not fully understood. In particular, models of human capital investment posit that workers trade-off expected costs and returns to determine their optimal level of training investments (e.g., Ben-Porath 1967; Sanders and Taber 2012). Given the uncertainty of future costs and benefits, workers' beliefs about returns to training are crucial determinants of their investment decisions. However, little is known about how these beliefs vary across workers and the extent to which they explain variation in firm training participation. ²

¹This chapter is based on co-authored work with Hanna Brosch and Philipp Lergetporer.

²In contrast, there is an extensive body of literature on schooling decisions that highlights the importance of students' beliefs about (non-)pecuniary returns to educational choices (see Giustinelli 2023 for an overview). However, evidence from schooling decisions cannot be readily transferred to

This paper investigates workers' beliefs about the (non-)pecuniary returns to firm training, and how they relate to training participation. We implemented an online survey in a representative sample of workers in Germany aged between 25 and 55 years ($N = 3701$) in 2024. To elicit return beliefs, we employ hypothetical investment scenarios (e.g., Dominitz and Manski 1996; Jensen 2010; Attanasio and Kaufmann 2014), asking workers to envision their careers six years later in two scenarios: one with annual firm training lasting two weeks and another without. We subsequently elicit workers' expected earnings in both scenarios, along with their subjective probabilities that various non-pecuniary career outcomes will materialize. For example, workers indicate how likely they believe it is to experience increased workplace stress, enjoy their work tasks, or quit their job. Our approach to eliciting subjective beliefs about firm training helps overcome a common identification problem in observational data. Specifically, observed training choices alone do not reveal the underlying decision process, as they are consistent with multiple combinations of preferences and beliefs. This makes it difficult to isolate the role of beliefs in determining training participation (Manski 2004).

In addition, we study whether differences in beliefs can explain lower firm training participation observed among lower-skilled compared to higher-skilled workers. Previous literature has established these differences (Bassanini et al. 2007; Lergetporer et al. 2023), but little is known about their determinants. This is despite the fact that replacing depreciated human capital is particularly important for lower-skilled workers, who are most vulnerable to adverse labor market outcomes due to skill-biased technological change (Goos et al. 2014).

Besides providing a descriptive analysis of existing skill gaps in training participation and beliefs, we also present experimental evidence on whether addressing training beliefs can influence gaps between lower-skilled and higher-skilled workers' intended training participation. Our light-touch intervention informs workers about the career benefits of training and the effectiveness of training for acquiring new skills, directly addressing known barriers to participation among lower-skilled workers (Osiander and Stephan 2018).

We start by documenting that workers perceive substantial positive career returns, both pecuniary and non-pecuniary, to participating in firm training. On average, workers expect their earnings six years later to increase by 8.62% through firm

firm training due to fundamental differences in the institutional settings, choice contexts, and the stakeholders involved (see Section 4.2 for a detailed discussion).

training. These return beliefs fall into the same ballpark as estimated earnings returns to firm training from observational data (e.g., Guo et al. 2024), indicating that workers' average beliefs are well-calibrated. However, there is considerable heterogeneity in these expectations, with nearly half of the sample expecting no returns. Additionally, workers expect substantial non-pecuniary career returns. Specifically, they expect an 8 percentage points higher probability of having a successful career with training than without, a 10 percentage points increase in promotion probability, and a 3 percentage points reduction in the risk of job loss. Beyond these career expectations, the probability of performing complex tasks is expected to increase by 14 percentage points through training, enjoyment of job tasks by 4 percentage points, and work-related stress by 2 percentage points. Differentiating by the type of skills that firm training targets (e.g., soft skills, IT, or administrative skills), we show that both pecuniary and non-pecuniary return beliefs are broadly similar across training contents. This suggests that workers' average expected returns to training are general in nature, rather than driven by specific types of training.

Investigating the predictive power of return beliefs for intended firm training participation, we show that most beliefs are significantly correlated with training intentions. This relationship persists even after controlling for a wide range of correlates of firm training participation, such as worker and firm characteristics, as well as detailed information about training characteristics. Lending support to the validity of our training-intentions measure, we find similar patterns when analyzing *actual* past training participation instead of intended future participation.

Next, we turn to differences between higher and lower-skilled workers. Descriptively, lower-skilled workers have much lower rates of both realized (67% vs. 46%) as well as intended (38% vs. 26%) firm training participation.³ In terms of beliefs, we do not find differences in expected relative earnings returns between groups. However, lower-skilled workers perceive lower non-pecuniary returns to firm training than higher-skilled workers. For example, while higher-skilled workers expect the probability of professional success to increase by almost 10 percentage points with firm training, this number amounts to merely 5.5 percentage points among lower-skilled workers. Strikingly, while lower-skilled workers barely

³The lower rate of intended compared to realized training participation is likely due to the fact that we explicitly elicited intended participation for training with a minimum duration of 40 hours, whereas no such threshold applied to realized participation.

expect the probability of enjoying workplace tasks to increase through firm training, higher-skilled workers expect returns of more than 7 percentage points in this dimension. They also expect their promotion probabilities to increase by 2.3 percentage points less than higher-skilled workers. Differences in beliefs account for approximately 12% of the “skill gap” in intended training participation.

These patterns suggests that beliefs are a major determinant of workers’ firm training decisions. To move beyond descriptive evidence, we next explore the causal relationship between worker beliefs and training intentions in a simple information-provision experiment (Haaland et al. 2023). In the experiment, we inform a randomly selected treatment group about the career benefits of training and its effectiveness for acquiring new skills, before eliciting training intentions as in the uninformed control group. We find that addressing worker beliefs through information provision significantly increases their training intentions, regardless of whether workers have to pay for part of the training costs themselves. Specifically, the information treatment significantly increases workers’ stated probability to participate in firm training the following year by more than two percentage points. Importantly, this effect is primarily driven by lower-skilled workers, whose treatment-induced increase in training intentions is about 3 percentage points – roughly double that of higher-skilled workers. Therefore, this proof-of-concept experiment shows that addressing workers’ beliefs about firm training can causally narrow the skill gap in training intentions.

We contribute to three strands of economic research. First, we add to the literature studying the determinants and consequences of firm training participation (see Black et al. 2023 for a comprehensive review). Regarding worker demographics, younger and more educated workers exhibit the highest training participation rates.⁴ Those in skill-intensive occupations also participate more frequently, implying sizable gaps between lower- and higher-skilled workers. Our survey data replicates these associations, affirming its reliability. Beyond demographics, Caliendo et al. (2020) and Caliendo et al. (2023) study how workers’ internal locus of control and risk tolerance relate to training participation, finding that both traits are positively associated with it. In contrast, empirical evidence is largely lacking on whether workers’ beliefs and expectations about firm training – which are key primitives in

⁴On the firm side, workers in larger and more innovative firms show higher participation rates (e.g., Bassanini et al. 2007). Recent papers show in addition that manager characteristics are an important determinant of training participation (Caliendo et al. 2024; Diaz et al. 2025; Caliendo et al. 2025).

models of human capital investment (Ben-Porath 1967; Sanders and Taber 2012) – influence training participation.⁵ We contribute to this literature by studying firm training from a new angle, specifically by examining a broad set of (non-)pecuniary return beliefs as potential determinants.

Turning to the consequences of firm training participation for workers, a related strand of research shows that training can increase productivity, promotion probabilities, and to a lesser extent, wages (Bartel 1995; Konings and Vanormelingen 2015; Adhvaryu et al. 2023). Additionally, Guo et al. (2024) demonstrate that training participation reduces workers' automation risk. Building on this, we extend the research by investigating workers' *beliefs* about such returns to firm training across various dimensions, and how beliefs influence training intentions and participation.

Second, our paper is related to the literature investigating the role of subjective expectations and beliefs as determinants of economically relevant life choices, particularly human capital investment decisions (e.g., Wiswall and Zafar 2021; see Giustinelli 2023 for an extensive review). This research largely shows that students' expectations about monetary returns and costs have only limited explanatory power for schooling and college choices (e.g., Wiswall and Zafar 2015; Boneva and Rauh 2021). Instead, non-monetary factors like the expected enjoyment of studying are the predominant drivers of educational decision and disparities across sociodemographic groups (e.g., Zafar 2013; Giustinelli 2016; Baker et al. 2018; Boneva and Rauh 2021). In contrast to this vast literature focusing on formal schooling and college choices, the role of workers' beliefs in explaining firm training decisions has not yet been studied. Importantly, insights on the determinants of schooling decisions are not readily transferable to firm training decisions, as they represent conceptually different types of human capital investment decisions, a distinction we elaborate on in Section 4.2. We address this research gap by directly measuring workers' beliefs about (non-)pecuniary returns to training and examining their impact on training intentions and participation.

Third, we contribute to a recent literature strand in labor economics that examines economic beliefs of workers and their relation to labor market behavior and outcomes.

⁵Recently, some papers have studied how workers' beliefs about automatability of their occupations influences their training intentions, generally finding a positive association (e.g., Innocenti and Golin 2022; Golin and Rauh 2022; Arntz et al. 2023; Lergetporer et al. 2023). Although these studies focus on beliefs within the context of firm training as we do, they address a different research question. Rather than examining beliefs about firm training directly, they investigate beliefs about the automatability of occupations, which may influence training participation.

Contrary to standard labor market models which assume that workers have accurate beliefs (e.g., Stigler 1961) several papers have documented that workers have imperfect knowledge of important labor market facts. For example, workers often hold incorrect beliefs about the expected duration of unemployment (Mueller et al. 2021), the wage distribution within their own firm (Cullen and Perez-Truglia 2022), or the external wage distribution that determines their outside options (Jäger et al. 2024).⁶ Evidence also shows that providing factual information to workers with imperfect knowledge can affect their decisions. For example, information about peers' wages or other applicants' wage expectations influences job search behavior (Card et al. 2012; Roussille 2024). Consequently, information provision that addresses worker beliefs can lead to adjustments of workers' expectations and their (intended) labor market behavior (Conlon et al. 2018; Belot et al. 2019; Jäger et al. 2024). We expand this literature towards the domain of firm training by studying workers' (non-)pecuniary return beliefs and their relationship to training intentions and behavior.

The remainder of this paper is structured as follows. Section 4.2 outlines our conceptual considerations and presents institutional background information on firm training in Germany. Section 4.3 presents our survey instrument. Section 4.4 presents our results, and Section 4.5 concludes.

4.2 Conceptual Considerations and Institutional Background

In this section, we first discuss the importance of firm training for a country's overall human-capital stock and how it differs from other forms of human capital investments (Section 4.2.1). Then, we provide an overview of the institutional framework governing firm training in Germany (Section 4.2.2).

4.2.1 Conceptual Considerations

Firm training, or on-the-job training provided by employers, is crucial for workers to acquire human capital, rivaling the importance of formal schooling at primary, secondary, and tertiary levels. The immense economic importance of firm training

⁶Caldwell et al. (2025b) show that workers' expected pay premia at firms other than their current employer are positively correlated with observed firm pay measures, including objective firm-specific pay premia.

is evident not only from its prevalence — with a large portion of the workforce in many countries participating annually (e.g., 66% in the U.S., or 60% in Germany; Black et al. 2023).⁷ The considerable resources that firms and workers invest in training, both in terms of direct costs and the opportunity costs of time, also indicate its substantial economic value-added. Indeed, firm training has been shown to boost productivity and wage growth for workers throughout their careers and to mitigate adverse effects of technological change (Konings and Vanormelingen 2015; Battisti et al. 2023). Similarly, firms benefit from firm training in multiple ways: it enhances productivity (Dearden et al. 2006; Bartel 1995; Adhvaryu et al. 2023), reduces absenteeism and turnover (Diaz et al. 2025), and frees up managerial resources as trained workers require less supervision (Espinosa and Stanton 2022). Additionally, firms commonly rely on training as a strategy to address widespread shortages of skilled labor (e.g., Baier et al. 2025). Surprisingly, given its economic significance, the body of economic research on firm training is relatively modest and is dwarfed by the much more extensive literature on all aspects of formal schooling – a point recently highlighted by Black et al. (2023).

The relative scarcity of research on firm training is *not* because findings from schooling research would be directly applicable to modeling workers' training decisions. Instead, fundamental conceptual differences between formal schooling and firm training demand independent investigation into firm-training decisions. A primary structural difference is that while formal primary and secondary schools typically share key features such as the duration and certificates awarded, firm training is markedly more heterogeneous. It is typically organized in a decentralized, less standardized manner, leading to variation in duration, content, or certification across training programs. This introduces challenges in defining and measuring firm training that are not encountered with formal schooling (see Black et al. 2023), an issue which we revisit when introducing our survey instruments in Section 4.3.

Second, from a choice-theoretic perspective, the framework governing workers' training decisions differs markedly from students' schooling choices. Training decisions typically recur throughout a worker's career, representing a series of interdependent, repeated choices. By contrast, schooling decisions, once made, generally establish the educational trajectory for several years. Moreover, because firm training inherently occurs while employed, the opportunity cost structure

⁷These percentages refer to workers who have engaged in informal firm training, which is training without formal certification, at least once over the past 12 months.

of these training decisions differs from that of schooling choices, as the next best alternative — continuing to work — is immediate and salient.⁸

A third difference between on-the-job training and formal schooling is the importance of firms in determining their workers' training participation. Firms play a central role in providing and financing training, as evidenced by the fact that they cover the cost of more than 90% of informal, non-certified training undertaken by employees across OECD countries (Black et al. 2023). This raises the question of how much influence firms versus workers have in the decision-making process regarding training participation. Traditionally, models of firm training treat it primarily as a decision for firms, emphasizing firm-specific returns to certain types of workers' skills (Becker 1962; Acemoglu and Pischke 1998; Acemoglu and Pischke 1999a; Autor 2001). When training enhances skills that are predominantly firm-specific and not easily transferable, firms have strong incentives to invest in training to realize productivity gains (Becker 1964; see Leuven 2005 for a review). However, firms have also been shown to invest in general skills of their workers. This investment behavior can be rationalized, for instance, through the presence of labor market frictions like wage compression (Acemoglu and Pischke 1999a; Pfeifer 2016). Consistently, 83% of workers in our survey report that the skills they acquired through firm training were at least partly transferable to other firms.

While firms play a crucial role in facilitating and funding training, the influence of workers in the decision-making process is also important, typically determining the actual uptake of training opportunities. Workers have economic incentives to acquire (general) human capital through training to benefit from the resulting productivity gains. Consistently, recent theories model firm training as joint decisions of firms and workers, with workers participating in training if it yields positive private returns (e.g., Caliendo et al. 2020). Our survey data underscore the role of worker preferences in shaping firm training participation: two thirds report having initiated their most recent firm training, either independently or in coordination with their

⁸Another form of human capital accumulation after schooling is learning-by-doing (see, e.g., Blandin 2018 and Arellano-Bover and Saltiel 2024). Unlike firm training, learning-by-doing does not involve foregone production, leading to different incentive structures – hence, we do not consider it in this paper. Skill acquisition in adulthood also takes place via government-sponsored active labor market programs to unemployed workers (e.g., Katz et al. 2022). For a summary of this literature, see McCall et al. (2016).

employer.⁹ This raises the central empirical question we address in this paper: How do workers' subjective beliefs about returns to training – theorized as a key determinant of training choices (Caliendo et al. 2020) – shape training decisions?¹⁰

4.2.2 Institutional Background: Firm Training in Germany

To set the stage, we first provide our definition of firm training and present key stylized facts about training in Germany. We establish these facts using the Adult Education Survey (AES), a key data source for measuring training participation across Europe. The AES provides standardized and comparable training measures, is conducted regularly, and is widely used by policymakers and researchers. It collects information on various forms of further training, including firm training (BMBF 2024). In Germany, the AES is conducted every two to three years. This probability-based survey collects data from a representative sample of the German population aged 18 to 69 (N = 9,820 in the latest 2022 wave).¹¹ It assesses firm training participation over the past 12 months, defined as *any training activity that occurs entirely or predominantly during working hours or where the employer covers direct training costs*. This includes both in-house training programs and external courses provided by third parties (BMBF 2024). We adopt this definition of firm training throughout this paper.¹²

Firm training differs from more formal types of training, such as technical degrees or apprenticeship programs, which also partly take place within firms. These programs are "classical" professional degrees, combining firm-based training

⁹Furthermore, one in four workers reports that a past training request was denied by their supervisor, highlighting workers' active pursuit of training opportunities – even if not always successful.

¹⁰Caliendo et al. 2020 argue theoretically that workers' return beliefs are a key determinant of training choices. However, they do not directly measure expected returns from training as we do, but instead proxy them with workers' locus of control. In contrast to the firm-training literature, vast empirical evidence highlights the importance of subjective return beliefs in shaping formal schooling choices (e.g., Giustinelli 2023)

¹¹Another possible data source, also used by Black et al. (2023), is the Programme for the International Assessment of Adult Competencies (PIAAC). However, we decided not to use PIAAC because its most recent data for Germany is from 2012.

¹²The AES also measures other forms of further training, namely individual job-related training (e.g., self-paid certificate courses or evening seminars), and non-occupational training (e.g., cooking classes or language courses for travel), as well as informal learning from colleagues or supervisors, which typically occurs in spontaneous, one-to-one interactions. We disregard these additional training types, as their greater heterogeneity in participation motives and barriers makes it harder to measure beliefs consistently. We focus on firm-provided training, as it is both the most common and economically most relevant form of further training.

with school-based education. They last more than six months and lead to an officially recognized qualification. As such, they mostly cater to younger individuals, making them less common across the entire workforce. In contrast, firm training is typically shorter in duration, often lasting just a single day and is designed to support ongoing skill development throughout the career, making it more widespread across all age groups (BMBF 2024).

According to the 2022 AES, 66% of employees in Germany reported participating in at least one firm training activity over the past twelve months.¹³ Substantial evidence shows that participation in firm training is heterogeneous across worker subgroups. In particular, workers in occupations demanding lower skills, older workers, those in precarious employment, and those holding jobs with high automatability are less likely to participate. Conversely, employees in the public sector, those in higher professional positions, with advanced education degrees, on permanent contracts, or in larger firms are more likely to participate. Participation rates are similar between men and women, between employees in eastern and western Germany, and between part-time and full-time workers (Heß et al. 2023; BMBF 2024). Reassuringly, these patterns of heterogeneity are largely reflected in our firm training measures, underlining the validity of our survey data (see Section 4.3.2).

To fix ideas, we next present some basic characteristics of firm training in the German context, specifically its duration, format, and content. Firm training courses are typically quite short: among employees surveyed in the AES, 47% of their courses span only a few hours. Average course duration is 23 hours, with a median duration of 8 hours. Only 9% of training activities exceed 40 hours.

The majority of firm training courses (71%) is conducted by the firms themselves, while the remaining share is provided by third-party institutions, such as chambers of commerce, industry associations, or external training providers (BMBF 2024). Unlike formal schooling, firm training tends to focus on applied, job-relevant skills. The content varies widely, ranging from automation technology courses for electricians to diversity training for managers and general digital skills training for office staff. A representative firm-level survey conducted in 2020 also reflects this

¹³This participation rate refers to employees aged 18-69 working full- or part-time. Our survey reports a comparable rate of 56% over the same period. Differences may arise from the more detailed measure in the AES, which includes up to 16 types of training activities and likely results in higher reporting compared to our shorter survey instrument. In addition, our sample is restricted to employees aged 25-55 and is slightly more educated than the general population (see Appendix Table 4.8).

diversity, showing that most training hours are invested in technical, practical, and job-specific skills, followed by training in customer orientation, general IT skills, problem-solving and teamwork skills ((Destatis) 2022).

4.3 The Survey

This section introduces our survey. We first outline our data collection process. Then, we explain how we measured firm training intentions and realizations, as well as workers' beliefs about returns to firm training. Finally, we describe our randomized information-provision experiment that tests how addressing worker beliefs can causally affect training intentions.

4.3.1 Data Collection

We conducted our online survey in May 2024 with a sample of 3,701 employed workers in Germany.¹⁴ Participants completed the survey on their own devices, with a median response time of 16 minutes. Due to our focus on human-capital investment decisions, we restricted our sampling to those aged between 25 and 55 years who work at least 15 hours per week.¹⁵ Respondents were recruited via the survey company Talk Online to represent average workers in Germany in terms of age, gender and state of residence.

Table 4.1 displays basic sample characteristics. Average age is 40.8 years, with 50% of the sample being female. On average, workers work 36.4 hours per week; 21% are employed part-time. Regarding professional degrees, 35% hold a vocational degree, 16% an advanced vocational degree¹⁶, and 45% a university degree. Workers report an average firm size of 453 employees, average tenure within the firm of 10.8 years, and average gross monthly earnings of 4,670 €.

In our analysis, we classify respondents as *higher-skilled* or *lower-skilled* based on a combination of their occupation's skill requirement level and their highest

¹⁴Online samples like ours have become a standard data source in economic research (Stantcheva 2023), and have been shown to represent the general population well (Grewenig et al. 2023).

¹⁵We disregard workers aged over 55 years because their impending retirement limits their incentives for human capital investments. For instance, Battisti et al. (2023) find that workers older than 55 fail to upgrade to more abstract jobs upon technological change. Similarly, individuals working less than 15 hours per week are not included in our sample due to their limited labor market attachment, which limits their firm training incentives and opportunities.

¹⁶Advanced vocational degrees include those from a technical school (*Fachschule*), master craftsman school (*Meisterschule*), technician school (*Technikerschule*), vocational academy (*Berufsakademie*), or professional academy (*Fachakademie*).

professional degree.¹⁷ Following Christoph et al. (2020), we classify workers as higher-skilled if their occupation mainly involves complex specialist tasks (skill level 3) or highly complex tasks (skill level 4), and they hold a technical school degree or a university degree. This classification applies to 46% of our sample. The remaining 54% are classified as lower-skilled. These are workers in occupations that mainly involve unskilled tasks (skill level 1) or professionally oriented tasks (skill level 2), or with at most a vocational degree.¹⁸ Appendix Table 4.12 provides a descriptive comparison of higher- and lower-skilled workers.

To gauge the representativeness of our data, Appendix Table 4.8 compares the characteristics of our sample (column 1) to the population of employees in Germany (column 2), using data from the German Federal Employment Agency (BA 2024). Confirming that our sampling strategy was successful, both datasets closely align in terms of the stratification variables age, gender and federal state of residence. Some differences emerge for occupational characteristics, which were not used for stratification. Specifically, our sample includes fewer workers in goods production, and more in business and IT compared to the population of German workers. Additionally, our sample includes fewer workers performing unskilled tasks, and has a higher share of full-time employment. While we do not claim representativeness along these dimensions, it is noteworthy that our diverse sample broadly captures the occupational diversity of the general workforce.

The survey structure is depicted in Appendix Figure 4.10. Respondents first provide basic demographic information and details about their job characteristics. We then elicit their intentions to participate in firm training in the future and their subjective beliefs about (non-)pecuniary returns to firm training. Next, we elicit their actual firm training participation over the past 12 months, a standard reference period for such questions. We subsequently conduct our randomized information-provision experiment to assess how addressing workers' beliefs can

¹⁷To measure occupations' skill requirement levels, we use the levels defined in the German Classification of Occupations 2010 (KldB 2010), which closely aligns with the International Standard Classification of Occupations (ISCO). To elicit these classifications, respondents selected their occupation from ten major groups and, if not listed, entered it in an open text field. After hand-coding the latter, occupational classifications are available for 99% of the sample.

¹⁸Note that workers in occupations with skill levels 3 or 4 may still be classified as lower-skilled if their most advanced professional degree is a vocational degree or less. This applies to 11% of our sample; 7% have skill level 3 and the remaining 4% skill level 4. Appendix 4.C presents robustness checks using an alternative classification, where the 4% of respondents in occupations requiring skill level 4 are classified as higher-skilled, regardless of their highest professional degree. Reassuringly, the results remain robust under this definition, although some of the differences between higher- and lower-skilled workers become somewhat smaller in absolute terms.

causally affect training intentions. Finally, we elicit respondents' risk and time preferences, along with their self-assessed response reliability (Dohmen and Jagelka 2024), to account for measurement error in our robustness analysis.

Table 4.1: Respondent characteristics

	Mean (1)	Std. Dev. (2)
Age	40.78	8.51
Female	0.50	0.50
Weekly hours	36.35	6.00
Weekly hours <35	0.21	0.40
Higher-skilled	0.46	0.50
Professional Degree		
Vocational training	0.35	0.48
Technical School	0.16	0.37
University	0.45	0.50
None/other	0.04	0.19
Number of workers at firm	452.95	411.84
Years at firm	10.78	9.20
Gross monthly earnings	4669.84	2553.91

Notes: Observations: N = 3701. Mean and Std. Dev. display averages and standard deviations. Gross monthly earnings are converted to full-time equivalent earnings and winsorized from below at the minimum wage (2,135 €) and from above at the 98th percentile of the German wage distribution (14,583 €).

4.3.2 Measuring Actual and Intended Firm Training Participation

To provide a comprehensive picture of how workers' beliefs shape training decisions, we measure both actual participation in the past 12 months and intended future participation. Self-reported survey questions like ours are widely recognized as the primary empirical method for measuring firm-training participation. Unlike formal schooling, firm training is a heterogeneous, non-standardized form of education: its content, duration, and level of formalization vary substantially across regions, firms, and individual workers. This heterogeneity makes accurate measurement of training activities through administrative data challenging. Worker surveys are therefore essential for capturing firm-training participation accurately (see Black et al. (2023) for an in-depth discussion of measurement issues).

Given the diverse ways in which firm training is organized, it is important to clearly define the concept before eliciting specific beliefs. We therefore provided

respondents with the following standardized definition. *Firm training refers to courses or events offered by your company to refresh existing professional skills or learn new ones. These are fully or partially financed by your company and can occur during or outside working hours. Firm training can be provided both externally and internally, ranging from a few hours to several months in duration.* This definition extends the AES definition by including information on training duration and provider type, reflecting the stylized facts presented in section 4.2.2, and is more detailed than definitions used in other surveys like the AES, PIAAC or SOEP. Given our focus on studying beliefs about firm training, this comparatively higher level of specificity ensures that respondents form their beliefs based on a clear and standardized understanding of the training concept.

Realized Firm Training. We ask respondents whether they participated in any firm training over the last 12 months. Our question format follows established surveys such as the AES and PIAAC. The exact question wording is provided in Appendix 4.F.1.

Firm Training Intentions. To measure firm training intentions, we use the self-reported probability (0–100%) of participating in at least one week (40 hours) of firm training over the next year. This duration reflects a realistic training scope, as the average realized training time among respondents who participated in training is 38 hours. Eliciting choice probabilities, rather than relying on binary responses or Likert scales, offers several advantages: it captures individual uncertainty, allows for interpersonal comparison, and conveys richer information (Manski 2004; Blass et al. 2010).

Table 4.2 presents actual and intended firm training participation in our sample. 56% of workers participated in firm training in the last 12 months. At the intensive margin, among those who participated, the average training duration was 38 hours, with a median of 16 hours. Nearly all training was employer-financed (98%) and conducted during working hours (96%). Regarding intended future participation, respondents report an average probability of 32% to engage in firm training in the next year. This lower figure compared to realized firm training likely reflects the higher duration threshold used in the intention measure, which specifies a minimum of 40 hours of training.

Different pieces of evidence support the validity of our firm training measures. First, realized and intended firm training participation in our sample follows patterns consistent with the Adult Education Survey (AES): younger, more educated

Table 4.2: Realized and intended firm training participation

	Mean (1)	Std. Dev. (2)
Participated in the last 12 months	55.53	49.70
Conditional on participation		
Hours in training	37.92	157.97
Training financed by employer	0.98	0.15
Training during working hours	0.96	0.19
Probability to participate in the next 12 months	31.57	32.62

Notes: Observations: $N = 3701$. Observations conditional on participation: $N = 2055$. Column 1 display averages and column 2 standard deviations. *Training financed by employer* equals 1 if financed at least partly by the employer; *Training during working hours* equals 1 if taking place at least partly during working hours.

workers and those in larger firms are more likely to participate, both retrospectively and prospectively (see Table 4.3). Conversely, participation rates are lower among workers with fewer working hours. Second, validating our measure of intended future training participation, we find that workers who participated in firm training in the past 12 months are 26 percentage points more likely to expect participating in the next year compared to those who did not. The correlation between intended and realized participation is 0.4. Given the path dependency of workers' training activities (see Pischke 2001), this correlation suggests that our expectations measure effectively captures future training participation. When presenting our results, we provide two further validations of our training measures. We show that (i) beliefs predict both realized and intended training participation (see Section 4.4), and (ii) our findings are robust to adjusting for reliability of responses (see Section 4.5).

4.3.3 Eliciting Workers' Beliefs

To elicit workers' beliefs about the (non-)pecuniary returns to training, we present each respondent with two hypothetical scenarios: one in which they participate in training lasting two weeks (80 hours) each calendar year until retirement, and one in which they do not participate in any training. For both scenarios, we ask them to imagine their professional lives in 2030 – six years after the survey. This approach – now standard in the literature on economic expectations (e.g., Dominitz and Manski 1996; Arcidiacono et al. 2020; Wiswall and Zafar 2021) – allows us to capture respondents' conditional expectations about future career outcomes

Table 4.3: Predictives of realized and intended firm training participation

	Realized firm training		Intended firm training	
	(1) Bivariate	(2) Multivariate	(3) Bivariate	(4) Multivariate
Age	-0.006*** (0.001)	-0.004*** (0.001)	-0.006*** (0.001)	-0.005*** (0.001)
Female	-0.079*** (0.016)	-0.038** (0.017)	-0.072*** (0.011)	-0.043*** (0.011)
Higher-skilled	0.216*** (0.016)	0.189*** (0.016)	0.124*** (0.011)	0.096*** (0.011)
Weekly hours <35	-0.066*** (0.020)	-0.012 (0.021)	-0.085*** (0.012)	-0.048*** (0.013)
Above median number of workers at firm	0.151*** (0.016)	0.131*** (0.016)	0.075*** (0.011)	0.058*** (0.010)
Observations		3701		3701
R ²		0.07		0.07

Notes: OLS regressions. Column 1 and 3: each cell represents the coefficient of a separate OLS regression. Column 2 and 4 depict one multivariate regression. Dependent variables: **Realized firm training**: dummy variable equal to 1 if firm training participation in the last 12 months, 0 otherwise. **Intended firm training**: Probability to participate in the next 12 months, ranging from 0 to 1. Robust standard errors (clustered at the individual level) in parentheses. Significance levels: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

under both scenarios. Moreover, it enables us to measure expectations about important non-pecuniary outcomes – such as job task complexity and enjoyment, or workplace stress – that are rarely available in administrative data. By focusing on within-individual variation in expectations, any differences between the two scenarios can be interpreted as the perceived impact of firm training.¹⁹

When designing the two scenarios, we made several deliberate choices to ensure they are both realistic and informative. First, we set the firm training duration at two weeks per year – twice the average annual training reported in our sample – to create a credible yet substantial contrast between scenarios. Second, we opted for a six-year time horizon when eliciting expectations to allow sufficient time for training returns to materialize.²⁰ Third, our scenarios contrast annual firm training with no training at all – both realistic scenarios given the path dependency in training participation (e.g., Pischke (2001)).

We elicit workers' beliefs about expected earnings and seven non-pecuniary career outcomes: whether they will have professional success, take on more complex job tasks, enjoy their job tasks, experience workplace stress, get promoted, quit

¹⁹In Section 4.5 we discuss the potential relevance of experimenter-demand effects in our design.

²⁰This duration falls within the typical time range used in labor or education economics research, where time horizons vary from weeks to decades depending on context (e.g., Wiswall and Zafar 2021; Mueller and Spinnewijn 2023).

their job, or be laid off. Figure 4.1 provides an overview of the belief elicitation task. We measure earnings beliefs in terms of gross monthly earnings, and ask respondents to report their beliefs about whether non-pecuniary outcomes materialize as probabilities (0-100%). For each outcome, respondents report two values – one for the scenario with firm training and one without.

To provide a more comprehensive picture of how return beliefs relate to training participation, we go beyond expected earnings returns and also elicit a broader set of non-pecuniary return dimensions. While expected earnings returns likely play a role in training decisions, workers may also care about factors such as job stability, career progression, or workplace stress. As Wiswall and Zafar (2021) highlight in the context of tertiary education, neglecting non-pecuniary return beliefs can lead to upward-biased estimates of how earnings expectations affect college major choice. Similarly, understanding the perceived costs and benefits of firm training requires measuring beliefs across multiple career outcomes. For many of the return dimensions, the sign of the expected returns is theoretically ambiguous (e.g., whether training increases or decreases workplace stress), and their relationship to actual and intended training participation is unclear a priori. Our survey is designed to address these empirical questions. In the following, we discuss the specific belief dimensions we consider in more detail.

Figure 4.1: Belief elicitation task

Expected outcomes in 2030	Scenarios	
	With firm training	Without firm training
Gross monthly earnings (in €)
<i>Probability (0–100%) to ...</i>		
... have professional success
... have more complex tasks than today
... enjoy tasks at work
... experience work-related stress
... get promoted at least once
... quit current job
... be laid off

Notes: This figure depicts the belief elicitation task in our survey. For screenshots of the exact version implemented in the survey, see Appendix Section 4.F.2.

Pecuniary Returns: Earnings Beliefs

To investigate perceived earnings returns to firm training, we elicit respondents' expected gross monthly earnings for the two hypothetical training scenarios. Respondents are instructed to assume that they will work the same number of hours as in their current job. This eliminates potential distortions in earnings expectations caused by expected labor supply changes over the next six years, including any such differences between the scenario with and without firm training. Additionally, respondents are told to disregard inflation, ensuring that expected earnings are not influenced by inflation expectations.

For our analysis, we convert these earnings expectations into full-time equivalents and winsorize the data at two thresholds: from below at the current full-time monthly gross minimum wage (2,135 €), and from above at the 98th percentile of the population wage distribution in Germany (14,583 €).²¹ We analyze both the absolute difference in expected earnings between the two scenarios, and the relative increase from the no-training to the training scenario. Focusing on relative differences has the advantage of accounting for baseline wage disparities – for example, between higher- and lower-skilled workers.

Beliefs about Non-Pecuniary Returns

In addition to earnings expectations, we elicit expectations about seven further career outcomes in 2030 to provide a more comprehensive picture of the perceived returns to firm training. These include one general outcome capturing whether workers expect to have a successful professional life, three job characteristics (i.e., taking on more complex tasks, enjoying tasks, and experiencing workplace stress), and three broader career outcomes (i.e., getting promoted, quitting job, or being laid off).²² For each of these seven items, respondents report the probability (0–100%) that the outcome will occur under both training scenarios.

The first outcome is a general question asking respondents to report the probability that they will have a successful professional life. This item serves as a catch-all measure to capture the overall expected effect of firm training on

²¹This adjustment affects 8% of the sample, with 5% of observations winsorized at the lower bound and 3% at the upper bound. Our qualitative results also hold when not winsorizing the data (see Appendix Section 4.D).

²²We refer to these belief measures as non-pecuniary returns to distinguish them from earnings expectations. However, we acknowledge that this distinction is not always clear-cut – for instance, promotions often coincide with changes in earnings.

workers' careers. The remaining sets of items are designed to break down this overall expectation into more specific beliefs about job characteristics and career outcomes.

The second set of outcomes captures beliefs about non-wage job characteristics by eliciting the expected probabilities of experiencing complex job tasks, enjoying tasks, and encountering work-related stress. The latter two dimensions reflect key aspects of job quality that workers are known to value (e.g., Maestas et al. 2023; Nagler et al. 2023). The nature of job tasks is a central determinant of labor market trajectories, with more complex tasks often linked to faster earnings growth (e.g., Acemoglu and Autor 2011; Deming 2024). Enjoyment of tasks also plays an important role, given strong associations between job satisfaction and productivity, retention, and absenteeism (Oswald et al. 2015; Böckerman and Ilmakunnas 2009; Krekel et al. 2019). Conversely, high levels of work-related stress can impair both physical and mental health, while also being associated with increased productivity (e.g., Nagler et al. 2023). A priori, it is unclear whether firm training improves or worsens the amenity value of job tasks – making the direction of its effects on (expected) job quality an empirical question. On the one hand, training may enhance job quality by equipping workers to take on more meaningful or satisfying tasks. On the other hand, it may also increase pressure and responsibility if newly acquired skills lead to more complex roles that some workers find overwhelming. For instance, Battisti et al. (2023) show that training yields routine-task workers to transition into more abstract jobs. Despite this, little is known about how workers themselves perceive changes in job characteristics resulting from firm training.

The third set of outcomes focuses on key career events typically examined in the labor-economics literature: the likelihood of receiving a promotion, quitting one's job, and being laid off. Promotions are widely recognized as key driver of earnings growth over the life cycle (Baker et al. 1994; Gibbons and Waldman 1999; Bronson and Thoursie 2019; Bayer and Kuhn 2023). If firm training enhances worker productivity – as documented in studies like Dearden et al. (2006) and Konings and Vanormelingen (2015) – then trained workers should be more likely to advance within their current firm (Benson et al. 2019). Since much of firm training builds general skills, training may also increase workers' promotion prospects at other employers. Workers' perceived probability of quitting captures their beliefs about outside options (e.g., Stigler 1961; Nelissen et al. 2017; Hoffman and Burks 2017; Jäger et al. 2024). Workers might expect that, despite training making them more productive, the associated wage increase within their firm may not fully reflect their

higher productivity – especially in firms with compressed wage structures (Pfeifer 2016). Workers may also believe that other firms pay differently even for similar roles, and that firm training could improve their chances of securing a better-paying job elsewhere (Caldwell et al. 2025b). Job loss, on the other hand, can result in substantial and persistent earnings losses for workers (Jacobson et al. 1993; Braxton and Taska 2023; Gulyas and Pytka 2025). Firm training may help mitigate this risk – for example, by enabling workers to transition away from tasks that are at risk of automation and toward more abstract, resilient job roles (Battisti et al. 2023).

Building on the existing literature on how training may affect non-pecuniary outcomes, we examine workers’ perceived returns along these dimensions – factors that should ultimately shape their training decisions.

4.3.4 Information-Provision Experiment

While the research interest of this paper is primarily descriptive – documenting the distribution of worker beliefs about firm training and their relationship to (intended) firm training participation – we also explore the extent to which addressing workers’ beliefs causally affect intended future training participation. To this end, we implement a simple information-provision experiment (Haaland et al. 2023) at the end of the survey. Importantly, the goal of this experiment is not to precisely pin down how each belief dimension affects training participation. Rather, it serves as a proof-of-concept to examine whether a generic information treatment that targets worker beliefs can affect training intentions.

The core of the experiment is a randomized information treatment targeting two key barriers to training participation identified in the literature: (1) workers’ uncertainty about the returns to firm training, and (2) workers’ beliefs that they are no longer accustomed to learning (Osiander and Stephan 2018).²³ Below, we describe the key design elements of the experiment. Appendix 4.F.3 provides the exact question wording.

Prior beliefs. First, we elicit respondents’ prior beliefs about those two barriers. In particular, we elicit respondents’ agreement with the following six statements: (i) firm training increases their salary in the long term, (ii) it improves their long-term career opportunities, (iii) they can improve at their job through practice, (iv) innate ability matters more than effort, (v) they would find it easy to learn new things

²³The experiment is preregistered in the AEA RCT (AEARCTR-0013319).

in training, and (vi) being out of the habit of learning is a barrier to training participation. In our analysis, we summarize items (i) and (ii) into a "positive return beliefs" index, and (iii) to (vi) into a "positive learning beliefs" index.

Randomized information provision. Subsequently, respondents in the treatment group are informed that research shows firm training increases wages and improves career opportunities. In addition, drawing on insights from positive mindset interventions (Yeager and Dweck 2020), they are told that anyone can learn new things through practice. Control group respondents do not receive any information.

Outcome: Probability of accepting a training offer. After the treatment, we elicit our outcomes of interest. Specifically, we ask respondents to imagine that their employer offers them the opportunity to participate in 80 hours of firm training during working hours. We then ask them to report the probability (0-100%) of accepting the offer in two different scenarios: (1) if the employer covers the full cost, and (2) if they must contribute 20% of their current net monthly wage to finance training costs. Varying the level of personal training costs helps mitigate potential ceiling effects. Unlike the earlier question eliciting training expectations, we now hold the supply of training (i.e., the employer's offer) fixed. This framing allows us to isolate respondents' willingness to accept a concrete training opportunity. We intentionally designed the question differently from the other questions on training intentions to minimize the risk of consistency bias (Falk and Zimmermann 2017).

4.4 Results

We present our empirical results in four steps. Section 4.4.1 documents beliefs about both pecuniary and non-pecuniary returns to firm training for the overall sample. Section 4.4.2 examines the extent to which these beliefs predict intended and realized training participation. Section 4.4.3 analyzes how return beliefs vary by skill level. Finally, Section 4.4.4 presents the results of our information-provision experiment, which tests whether beliefs about firm training are malleable.

4.4.1 Beliefs about Returns to Firm Training

We begin by analyzing workers' beliefs about pecuniary returns – that is, expected wage differences with and without firm training (Section 4.4.1), and then turn to beliefs about non-pecuniary returns (Section 4.4.1).

Beliefs about Pecuniary Returns

On average, workers expect gross monthly earnings of 5,481 € in 2030 in the scenario with firm training, compared to 5,140 € without training (see column 1 of Table 4.25). The absolute difference of 341 € corresponds to an expected relative earnings return of 8.62%. These perceived returns are substantial, roughly aligning with empirical estimates of the returns to one additional year of full-time schooling (e.g., Hanushek et al. 2015).²⁴ As expected, returns are lower when focusing on median beliefs: respondents report median expected earnings of 4,686 € with training and 4,375 € without, implying a median expected return of 2.04%. This pattern suggests that (i) the average expected returns reported in column 1 are largely driven by workers with relatively high earnings expectations, and (ii) earnings beliefs are fairly dispersed. This dispersion is also reflected in the fact that only about half of our sample (52%) of respondents expect any positive return (i.e., higher earnings with training than without), 46% expect zero returns, and only 3% expect negative returns.

While actual returns to training provide an objective benchmark, it is individuals' subjective return beliefs that should ultimately determine their training decisions. Next, we compare these beliefs to empirical return estimates to assess how well expectations align with realized outcomes. Using German PIAAC data, Guo et al. (2024) find that participation in job training leads to sizeable wage gains. In their preferred specification, they estimate an average wage increase of estimate earnings returns to training of 12.6% for participating in training in the previous year.²⁵ In Section 4.4.3, we show that workers' return beliefs are also broadly consistent with empirical estimates when differentiating between lower- and higher-skilled workers. Thus, workers' beliefs are, on average, reasonably well calibrated.

Expected returns to training likely depend on the specific content of the training and the labor-market-relevant skills it imparts. Next, we therefore differentiate between specific training contents. To this end, we asked respondents for which

²⁴While sizable, expected earnings returns to firm training are somewhat smaller than those documented in other educational domains. For example, Wiswall and Zafar (2021) find that female students expect a 30% earnings increase at age 23 from completing a science or business degree instead of a humanities degree. Boneva et al. (2022) report that undergraduate students in England expect a 15% earnings premium at age 35 from completing a postgraduate degree compared to holding only an undergraduate degree. However, these scenarios comprise different types of education in multiple contexts which also differ in the amount of hours, making return beliefs hard to compare.

²⁵Their definition of job training includes both firm training and self-initiated job-related training undertaken without any employer support.

types of skills their employer provides an opportunity to improve them through training: (1) soft skills and personal development, (2) IT skills, (3) office and administrative skills, and (4) technical, practical, or job-specific skills.²⁶ Figure 4.6 presents average relative expected return beliefs separately for respondents with access to different types of training. Regarding earnings expectations, we find that all categories ((1)–(4)) are associated with substantial expected returns, with IT skills showing somewhat higher perceived returns than the others. This suggests that the documented positive average expected earnings return is not driven by outliers or niche training types, but rather reflect broadly shared perceptions across various types of skill training.

Table 4.4: Earnings expectations descriptives

	Mean (1)	Std. Dev. (2)	Median (3)
Expected earnings in €			
With firm training	5481	2987	4686
Without firm training	5140	2898	4375
Individual expected returns			
Absolute	341	867	100
Relative in %	8.62	29.43	2.04
Share positive	0.51	0.50	1

Notes: $N = 3701$. Means and standard deviations refer to worker-reported expectations of monthly gross earnings. *Individual expected returns* contain summary statistics of workers' perceived absolute and relative returns. Relative returns correspond to the within-worker difference in expected earnings across scenarios divided by earnings in the scenario without firm training. All figures are in full-time equivalents. Earnings expectations are winsorized from below at the minimum wage (2,135€) and from above at the 98th percentile of the German wage distribution (14,583€).

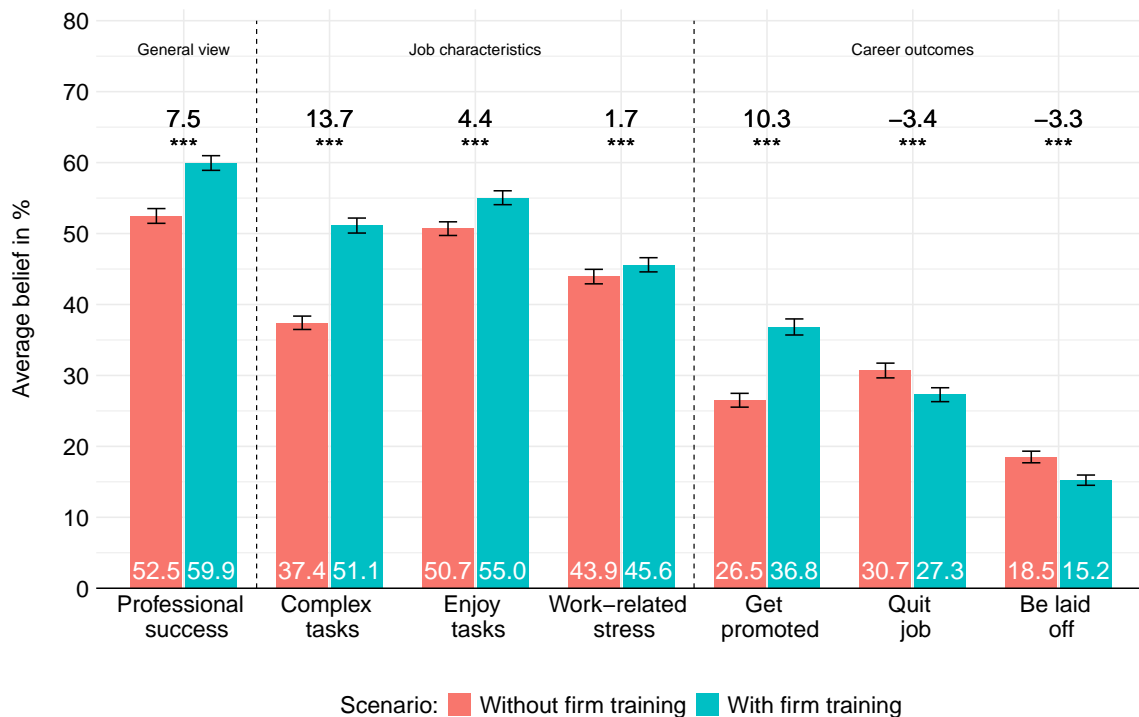
Beliefs about Non-Pecuniary Returns

Figure 4.2 presents average expectations about non-pecuniary career outcomes in 2030 under the scenario with firm training (blue bars) and without firm training

²⁶For this multiple choice question, the exact wording was as follows: "For which types of skills does your employer offer you the opportunity to learn or improve them through firm training?" Next to the skill types mentioned above, the answer options included "none", and "other". 22% of respondents selected "none", 30% selected one option, and 47% selected two to four options. As a result, 22% of respondents are not represented in the figure, while 47% appear in multiple categories.

(red bars), along with the difference between them (indicated above the bars). On average, respondents expect a higher likelihood of professional success with firm training (59.9%) than without it (52.5%), corresponding to an expected return of over 7 percentage points. Thus, workers perceive firm training as beneficial for their professional life – an assessment consistent with prior evidence on the benefits of training programs (Black et al. 2023).

Figure 4.2: Beliefs about non-pecuniary returns



Notes: This figure displays average probabilities for each event under two scenarios: with and without firm training. Bars show means and 95% confidence intervals. The number above each pair of bars indicates the mean difference between the two scenarios and stars denote significance levels from paired t-tests. Significance levels: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

When analyzing the perceived effect of firm training on expected job characteristics six years later, workers anticipate a 13.7 percentage points higher likelihood of performing more complex tasks in the scenario with firm training (51.1%) compared to the scenario without (37.5%).²⁷ The expected probability that workers will enjoy

²⁷The fact that a substantial share also expects more complex tasks without firm training aligns with, for example, Werner et al. (2022), who show that a plurality of German citizens expect job tasks to become more demanding in the future.

their job tasks is also slightly higher with firm training (55.0%) than without (50.7%), implying a modest positive difference of 4.4 percentage points.

In contrast, work-related stress is expected to remain largely unchanged across scenarios: respondents report a 45.6% likelihood to encounter work-related stress with training and 43.9% without, a small difference of 1.7 percentage points. This suggests that workers do not perceive firm training to substantially increase stress levels — alleviating concerns that more complex tasks might come at the cost of higher psychological strain.

In terms of broader career outcomes, workers expect a substantially higher likelihood of getting promoted in the scenario with firm training (36.8%), compared to the scenario without (26.5%), implying a substantial positive expected return of 10.3 percentage points. This sizable gap highlights workers' belief in the role of training in career advancement. The expected likelihood of quitting is 3.4 percentage points lower with firm training (27.3%) than without (30.7%), and the expected risk of being laid off decreases by 3.3 percentage points with firm training (15.2% versus 18.5%).

Figure 4.6 displays these non-pecuniary return beliefs broken down by the type of training content respondents have access to. As with pecuniary return beliefs, we find no pronounced differences across training types. The signs of the expected effects are consistent, and the magnitudes are broadly similar across all skill categories. This suggests that perceived non-pecuniary returns to firm training are largely universal, regardless of training content.

In summary, workers perceive firm training to positively affect not only their earnings but also key non-pecuniary career outcomes, with the largest perceived effects related to promotion opportunities and job-task complexity. Importantly, these effects are not expected to come at the cost of increased stress, implying an overall positive perception of firm training. These findings highlight that non-pecuniary return beliefs may play an important role in shaping training decisions. In the next section, we examine how (non-)pecuniary beliefs relate to actual and intended training participation.

4.4.2 Return Beliefs Predict Intended and Realized Training Participation

Having documented that workers expect substantial (non-)pecuniary returns to firm training, we next investigate how these expectations relate to intended and realized firm training participation.

We begin by analyzing intended future training participation. Specifically, we estimate worker-level regressions in which the dependent variable is the self-reported probability of participating in at least one week (40 hours) of firm training in the next year. The key independent variables are workers' return beliefs – measured as the percentage point difference in expected (non-)pecuniary returns between the scenario with and without firm training.

Figure 4.3 shows that several dimensions of return beliefs are significantly associated with intended firm training participation. In particular, expected returns regarding earnings, professional success, task complexity, task enjoyment, and promotion prospects show positive and significant coefficients. Conversely, expected differences in the likelihood of quitting one's job or being laid off across training scenarios are negatively related to training intentions. By contrast, expected effects of training on work-related stress are not significantly associated with training intentions. The magnitude of the regression coefficients varies notably across belief domains. Without additional controls, a one percentage point increase in expected earnings returns is associated with a 0.09 percentage point increase in intended training participation (see pink coefficients). In contrast, non-pecuniary return beliefs – such as those about professional success and task enjoyment – show substantially larger coefficients (0.23 and 0.22 percentage points, respectively). This pattern suggests that non-pecuniary return beliefs are particularly strong predictors of training intentions, underscoring the importance of considering a broader set of expected returns beyond earnings alone.

These general patterns remain robust when we control for a rich set of worker, firm, and past training characteristics, as well as economic-sector fixed effects (see brown coefficients). While the coefficients naturally attenuate, all previously significant associations remain statistically significant after conditioning on these variables.

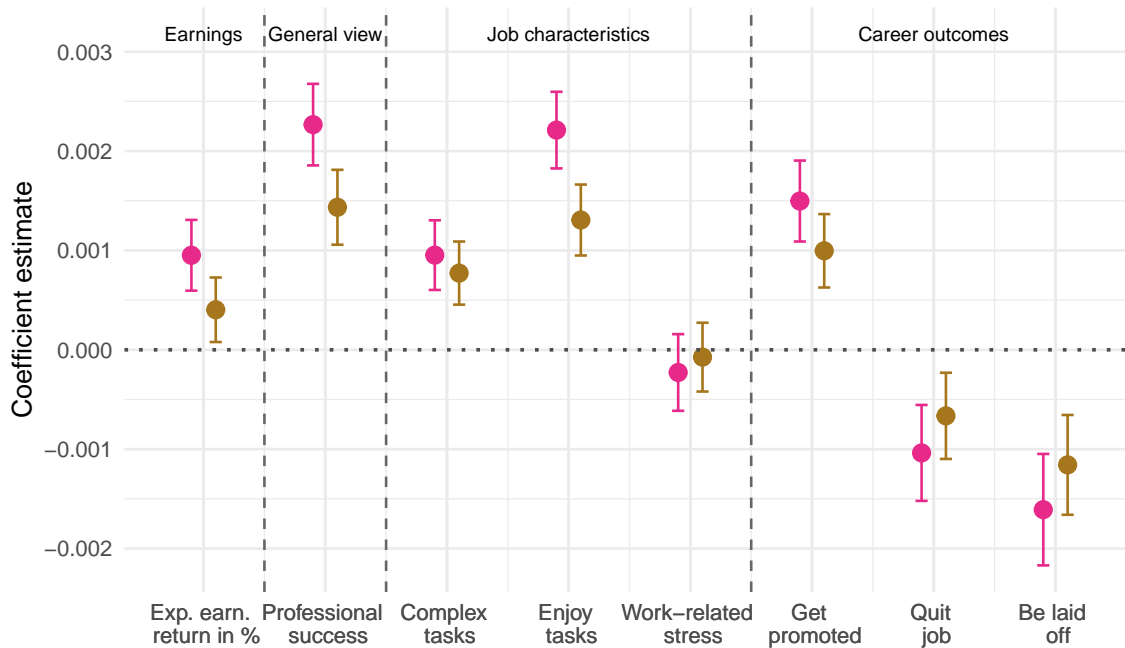
A common criticism against self-reported intentions is that they may not always translate into actual behavior. This concern also applies to our measure of intended future training participation. To address this, we complement our analysis by

examining how workers' return beliefs relate to realized training behavior. Using the same regression model as before, Appendix Figure presents results with the dependent variable indicating whether a respondent participated in any firm training in the past 12 months. The overall pattern closely mirrors that for intended training: without controls, both earnings and – especially – non-pecuniary return beliefs, such as those about professional success and task enjoyment, strongly predict realized training participation. While effect sizes are somewhat attenuated with controls, the general pattern remains intact.²⁸

In sum, our analysis shows that workers' beliefs about the returns to firm training are strong predictors of actual training participation, with beliefs about non-pecuniary career outcomes playing an especially important role.

²⁸Naturally, explaining past training participation with current return beliefs, as we do, has its own limitations. Most notably, beliefs may evolve in response to past training, leading to potential reverse causality. While we do not claim a causal interpretation of our findings, it is reassuring that results are highly consistent for both intended and realized training, reinforcing the link between return beliefs and training behavior.

Figure 4.3: Do return beliefs predict intended firm training participation?



Controls ● None ● Worker-, firm-, and training characteristics, and sector fixed effects

Notes: Coefficient estimates from a regression of the probability to participate in firm training next year on the expected earnings return in % (0-100) in the leftmost column and on the difference in values between the two hypothetical scenarios in percentage points for the remaining non-pecuniary belief dimensions. "None" indicates point estimates from a bivariate regression, while "Worker, firm, and detailed training characteristics + sector FE" indicates point estimates upon including these controls. Worker and firm characteristics: workers' age, gender, risk aversion, patience, locus of control, and current full-time equivalent earnings, as well as dummies indicating part-time work and above median firm size. Detailed training characteristics: indicators for the absence of training opportunities at the current workplace, whether workers were denied training previously, whether workers think their job is more demanding than a few years ago, and whether it is possible to accumulate skills through learning-by-doing. Fixed effects for five sectors are included.

4.4.3 Belief Gaps Between Lower- and Higher-Skilled Workers

Motivated by evidence that lower-skilled workers have lower levels of both actual and intended participation in firm training (e.g., Bassanini et al. 2007; Lergetporer et al. 2023), we now turn to differences in return beliefs by skill level. We first document differences in pecuniary and non-pecuniary return beliefs between higher- and lower-skilled workers. Then, we assess how much of the gap in intended training participation these beliefs can account for.

Beliefs about Pecuniary Returns by Workers' Skill Level

Table 4.26 reports expected gross monthly earnings with and without firm training, along with the resulting (relative) return beliefs, separately for lower- and higher-skilled workers.²⁹ On average, lower-skilled workers expect substantially lower earnings than higher-skilled workers – both in the training scenario (4,571 € vs. 6,543 €) and without training (4,275 € vs. 6,151 €) – which broadly reflects known empirical earnings gaps (see Appendix Table 4.11 for a comparison of earnings and other characteristics between lower- and higher-skilled workers). While absolute expected returns to firm training are somewhat lower for lower-skilled workers (297 € vs. 392 €), relative expected returns are nearly identical across groups (8.45% vs. 8.82%). Interestingly, a slightly higher share of lower-skilled workers (54%) expect positive earnings returns compared to higher-skilled workers (49%). Overall, these patterns suggest that beliefs about the (relative) pecuniary returns to firm training are quite similar across skill groups.

Next, we compare these expected earnings returns to firm training with empirical returns for lower- and higher-skilled workers. Based on PIAAC data for Germany, the average return to firm training participation is 12.6%, with no significant heterogeneity between lower- and higher-skilled workers.³⁰ Despite being larger, these estimates fall into the same ballpark as our elicited return expectations, suggesting that neither group holds major average misperceptions about the pecuniary returns to firm training.³¹

Beliefs about Non-Pecuniary Returns by Workers' Skill Level

We now turn to differences in beliefs about non-pecuniary returns to firm training across skill groups. Figure 4.4 depicts return beliefs of lower-skilled (green bars) and higher-skilled workers (orange bars). As in previous sections, we define beliefs about non-pecuniary return as the difference in the stated likelihood that a given career outcome will occur in 2030 between the scenarios with and without firm training.

²⁹We classify workers as lower- and higher-skilled based on their occupation's required skill level and their highest completed professional degree. For details, see Section 4.3.1.

³⁰We thank Yuchen Guo for calculating these figures using the German data as well as our sample restrictions and classification of higher- and lower-skilled workers, based on data from Guo et al. (2024).

³¹Potential reasons why the estimates differ are manifold. One is that the definition of firm training used in PIAAC differs slightly from the one used in this paper. In particular, the PIAAC definition includes training that is job-related but does not qualify as organized on-the-job training.

Table 4.5: Earnings expectations by workers' skill-level

	Higher-skilled		Lower-skilled		Difference Lower – Higher (5)
	Mean (1)	Std. Dev. (2)	Mean (3)	Std. Dev. (4)	
Expected earnings in €					
With firm training	6,543	3,353	4,571	2,268	-1,971***
Without firm training	6,151	3,297	4,275	2,158	-1,876***
Individual expected returns					
Absolute	392	949	297	788	-95***
Relative in %	8.82	30.71	8.45	28.29	-0.37
Share positive	0.48	0.50	0.53	0.50	0.05***
Observations		1,993		1,708	3,701

Notes: Mean and Std. Dev. display averages and standard deviations of worker-reported expected monthly gross earnings by workers' skill-level. Individual expected returns average perceived individual absolute and relative returns across workers. Relative returns correspond to absolute returns divided by earnings in the scenario without firm training. "Lower – Higher" is the mean difference. Earnings expectations are converted to full-time equivalents, winsorized from below at the minimum wage (2,135€) and from above at the 98th percentile of the German wage distribution (14,583€), and then converted to full-time equivalent earnings. Significance: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

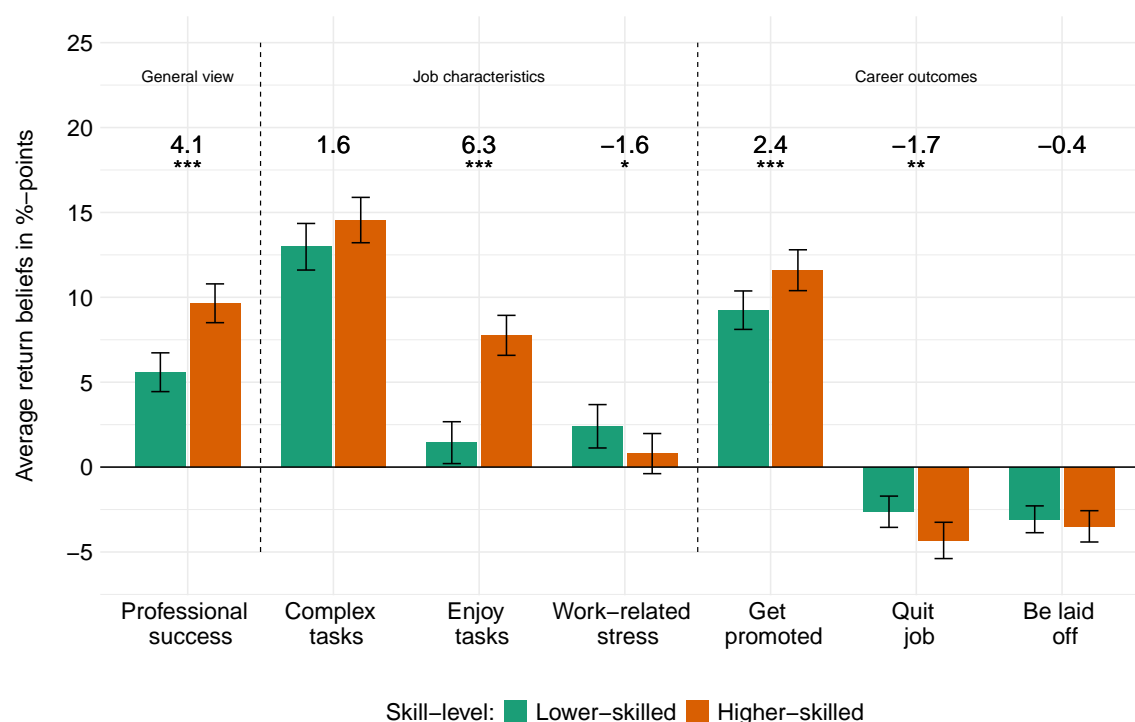
The general pattern across these career outcomes is that higher-skilled workers consistently perceive firm training as having more positive effects. On average, they expect training to increase their likelihood of professional success by nearly 10 percentage points, compared to only 5.6 percentage points among lower-skilled workers.

Regarding job characteristics, the expected impact of firm training on task complexity and work-related stress is relatively similar across skill groups. However, notable differences emerge in beliefs about task enjoyment, with higher-skilled workers expecting substantially larger gains from training (7.8 percentage points) than lower-skilled workers (1.4 percentage points). More positive expectations among higher-skilled workers also extend to broader career outcomes. They anticipate a larger positive effect of training on their promotion prospects and

a stronger negative effect on the likelihood of quitting their job. In contrast, expectations regarding the impact of training on the probability of being laid off are similar across skill groups.

In sum, these findings indicate that higher-skilled workers expect greater returns to firm training – particularly in the non-pecuniary domain – which may help explain their higher participation rates.³²

Figure 4.4: Difference in return beliefs by workers' skill-level



Notes: This figure displays average belief gaps between lower- and higher-skilled individuals for each event. Bars show group means with 95% confidence intervals. The number above each pair of bars indicates the mean difference between groups and stars denote significance levels from independent t-tests. Significance levels: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

Can Beliefs Explain Skill Gaps in Firm Training Participation?

In the previous sections, we documented that higher-skilled workers hold more optimistic beliefs about the returns to firm training than lower-skilled workers. We now ask to what extent these belief gaps help explain the observed skill gap in (intended) training participation.

³²Appendix Table 4.10 probes the robustness of the skilled coefficient by including controls....

Table 4.6 regresses intended firm training participation on a higher-skilled dummy. Column 1 shows that higher-skilled workers are 12.4 percentage points more likely to intend to participate in training next year. Controlling for return beliefs reduces this gap to 10.9 percentage points (column 2), implying that beliefs explain roughly 12% of the skill gap. A similar pattern holds when controlling for a rich set of worker, firm and training characteristics and occupation fixed effects: with these covariates, the skill gap reduces to 3.5 percentage points (column 3). Additionally accounting for worker beliefs reduces the gap to 3.2 percentage points, implying a reduction of about 9%.³³

In sum, accounting for beliefs explains a meaningful share of the skill gap in training intentions. Naturally, this descriptive analysis does not speak to whether beliefs are a *causal* determinant of the gap. To add causal structure to this relationship, the next section presents results from our information-provision experiment, examining whether addressing beliefs can reduce the skill gap in training intentions.

Table 4.6: Skill gap in firm training intentions

Dep. Var.: Training intentions	(1)	(2)	(3)	(4)
Higher-skilled	12.396*** (1.056)	10.897*** (1.043)	3.516*** (1.064)	3.166*** (1.057)
Beliefs		Yes		Yes
Worker and firm char.			Yes	Yes
Training char.			Yes	Yes
Sector			Yes	Yes
R^2	0.036	0.079	0.214	0.233

Notes: $N = 3,701$. Coefficient estimates of a dummy indicating whether a worker is higher skilled in a regression of the probability of participating in firm training next year on varying sets of controls. Beliefs: belief gaps in pecuniary and non-pecuniary dimensions. Worker and firm char.: workers' age, gender, risk aversion, patience, locus of control, and current full-time equivalent earnings, plus dummies for part-time work and above-median firm size. Training char.: indicators for the absence of training opportunities at the current workplace, whether workers were denied training previously, whether the job is more demanding than a few years ago, and whether skills can be accumulated through learning-by-doing. Sector: fixed effects for five sectors. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

³³The pattern is similar for realized training behavior over the past year, though the share explained by beliefs is somewhat smaller (see Appendix Table 4.14).

4.4.4 Experimental Results

This section presents the results from our information-provision experiment, which serves as a proof-of-concept to test whether addressing worker beliefs can causally reduce the skill gap in training intentions. The balancing tests in Appendix Table 4.15 confirm that the random assignment of treatment and control groups worked as intended: only two out of twelve pairwise comparisons of worker characteristics between experimental groups yield marginally significant differences ($p < 0.1$), as would be approximately expected by pure chance.

Table 4.7 presents the main experimental results. It shows the effects of the information treatment — which informed respondents that firm training improves wages and career opportunities and that anyone can learn new things through practice — on respondents' reported probability of accepting a firm training offer. Columns 1 and 2 display results for a scenario in which the employer covers the full cost of training, while columns 3 and 4 refer to a scenario in which workers contribute 20% of their current net monthly salary. Odd-numbered columns report the average treatment effect, and even-numbered columns show heterogeneous treatment effects by skill-level.

Descriptively, the average intention to participate in firm training in the control group is 81% when training is free of charge and drops to 31% when workers are required to contribute to the costs (see control mean). The substantial gap between training intentions in the two scenarios highlights the critical role of personal costs in shaping training decisions. Moreover, in line with earlier results, higher-skilled individuals consistently report greater intentions to participate in training across both scenarios.

The information treatment increases workers' willingness to participate in firm training by 2.2 percentage points, both when training is free and when workers are required to contribute to the costs. Notably, the effects are concentrated among lower-skilled workers, who show significant treatment effects of 3.0 and 2.8 percentage points in the two scenarios, respectively. In contrast, effects for higher-skilled workers are smaller and not statistically significant. Lower-skilled workers, who generally hold more pessimistic beliefs about the returns to firm training, are more responsive to an intervention that addresses related beliefs. This greater responsiveness helps narrow the skill gap in training intentions.

Table 4.7: Information treatment effects on training intentions

	Probability of accepting a training offer			
	.. with no costs		.. with costs	
	(1)	(2)	(3)	(4)
Treatment	2.234** (0.896)	3.034** (1.306)	2.138** (1.017)	2.847** (1.354)
Treatment x Higher-skilled		-1.667 (1.789)		-1.388 (2.038)
Higher-skilled		3.581*** (1.306)		6.843*** (1.495)
Treatment effect on higher-skilled		1.367 (1.212)		1.459 (1.523)
Control mean	81.03	81.03	30.70	30.70
Observations	3701	3701	3701	3701
Controls	✓	✓	✓	✓

Notes: OLS regressions. Dependent variable: *Probability of accepting a training offer* of 80 hours of the employer *with no costs*: the employer fully covers the costs, or *with costs*: 20% of their current net monthly salary must be paid by themselves. Robust standard errors (clustered at the individual level) in parentheses. Significance of treatment effect for higher-skilled respondents are calculated using post-estimation Wald tests. Control mean: mean of the outcome variable in the control group. Controls include age, dummies for female and ten occupational major groups (KldB 2010), risk preferences and patience. Significance levels: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

Next, we study heterogeneous treatment effects by prior beliefs.³⁴ The information treatment targets two key barriers to firm training participation: pessimistic beliefs about expected returns and perceived learning barriers. Before the treatment, we elicited various beliefs related to both barriers. We use answers to these individual prior-belief questions to construct summary indices "positive belief returns" and "positive learning returns" as described in Section 3.4 (see Appendix Table 4.16 for answers to the individual belief questions). The heterogeneity analysis in Appendix Table 4.17 shows that workers with more pessimistic views – both about training returns and learning barriers – respond more strongly to the information treatment than those with more optimistic beliefs. This suggests that information interventions hold particular promise for encouraging participation among workers

³⁴We also preregistered heterogeneity analyses by age. The information treatment significantly increases training intentions for both younger and older workers, with younger workers showing larger effects (see Appendix Table 4.18).

who are most discouraged from training.

In sum, our experimental results demonstrate that addressing workers' beliefs can influence their training intentions. While the experiment was designed as a general proof of concept – rather than to isolate the effects of specific belief dimensions – we view this as a promising direction for future research. Furthermore, it would be interesting to study in a field experiment whether changes in training intentions induced by information treatments translate into actual participation.

4.5 Discussion and Conclusion

This paper investigates workers' beliefs about the returns to firm training using survey data from over 3,500 respondents in Germany. We show that workers expect substantial returns – not only in terms of earnings, but also across a range of non-pecuniary career outcomes such as professional success, task enjoyment, and promotion prospects. These beliefs strongly predict both actual and intended participation in training. Lower-skilled workers expect lower returns than their higher-skilled counterparts and are less likely to participate in training – a gap that is partly explained by differences in return beliefs. Moving beyond descriptive findings, we also provide experimental evidence that addressing such beliefs through information provision increases training intentions, particularly among lower-skilled workers.

While survey data like ours facilitate observing beliefs and expectations – which are usually not available in administrative or other data sources – they also come with limitations. First, for many belief dimensions, no objective benchmark exists to assess whether beliefs are accurate, which is important to consider when designing targeted information interventions. While we can compare earnings return beliefs to empirical estimates and find them generally well-calibrated, such validation is not possible for non-pecuniary beliefs due to the lack of objective benchmarks. That said, our primary interest lies in subjective beliefs, which ultimately shape behavior regardless of objective accuracy.

Second, one of our key outcomes is intended future behavior (i.e., training participation), which is sometimes criticized for uncertain predictive validity. Reassuringly, our findings on the role of return beliefs are robust when using realized rather than intended participation, suggesting that our results are not driven by the choice of outcome measure.

Third, our study is primarily descriptive, and while it establishes strong associations between return beliefs and training behavior, it is not designed to fully disentangle underlying causal mechanisms. Our information-provision experiment provides first evidence that addressing beliefs can causally shift training intentions, especially among lower-skilled workers. However, it does not isolate the effects of specific belief dimensions—an important task for future research.

Fourth, eliciting expectations under different scenarios within the same respondent (e.g., Dominitz and Manski 1996; Wiswall and Zafar 2021) raises potential concerns about experimenter-demand effects. Since differences across scenarios are salient by design, respondents may feel obliged to report responses they think align with the experimenter’s expectations rather than their own views (de Quidt et al. 2018). While this is a theoretical concern, we consider it unlikely in our context: evidence suggests that demand effects are generally small in survey experiments (Mummolo and Peterson 2019), and the incentive to please an experimenter in an anonymous online setting is minimal. Overall, we believe the advantages of a within-subject design – in terms of statistical power and richness of individual-level data (List 2025) – outweigh the potential downsides stemming from stronger identification assumptions compared to between-subject designs.

Finally, a potential concern specific to survey data is that respondents may lack the self-knowledge required to accurately assess how firm training would affect their future outcomes. Even when respondents exert effort in answering, they may still face cognitive uncertainty – i.e., uncertainty about their own preferences (Enke and Graeber 2023; Dohmen and Jagelka 2024). This form of uncertainty can introduce residual measurement error that differs from inattentiveness or satisficing. It may lead some respondents to under- or overestimate returns to firm training, particularly in non-pecuniary dimensions that are inherently harder to verify, and thus distort their stated training intentions. While we cannot fully rule out this concern, we take a direct step to address it: Following Dohmen and Jagelka (2024), we elicited all respondents’ self-reported answer reliability as a proxy for cognitive certainty.³⁵ In Appendix 4.E, we replicate our main results for the subsample of respondents with reliable answers. Results remain virtually unchanged, suggesting that limited self-knowledge does not drive our findings.

³⁵We follow Dohmen and Jagelka (2024) and use their general reliability item: “Please indicate on the scale below how reliable your answers to this survey are,” measured on an 11-point Likert scale from 0 to 10. Respondents are classified as “reliable” if they select 9 or 10. In our sample, 80% fall into this category, compared to 70% in Dohmen and Jagelka’s study.

From a policy perspective, our findings suggest that beliefs about firm training play a critical role in shaping training participation. Leveraging these beliefs may be a viable strategy to stimulate upskilling, particularly among lower-skilled workers. Our experimental evidence shows that addressing belief-based barriers can causally increase training intentions in this group. Targeted information and encouragement campaigns could therefore be a promising tool for policymakers to engage more vulnerable workers in training programs – ultimately supporting their long-term labor market prospects.

4.A Appendix A: Additional Figures and Tables

Table 4.8: Representativeness

	Sample	Admin data	(1)-(2)
	(1)	(2)	(3)
Age groups			
25-34	28.42	32.76	-4.34
35-44	35.82	35.12	0.70
45-55	35.76	32.12	3.64
Gender			
Male	50.32	53.63	-3.31
Female	49.68	46.37	3.31
Federal state			
Baden-Wuerttemberg	12.99	14.07	-1.08
Bavaria	16.02	17.28	-1.26
Berlin	4.11	5.19	-1.08
Brandenburg	3.13	0.25	2.88
Bremen	0.76	0.10	0.66
Hamburg	2.24	0.33	1.91
Hesse	7.62	8.08	-0.46
Mecklenburg-Western Pomerania	1.76	0.16	1.60
Lower Saxony	10.08	8.86	1.22
North Rhine-Westphalia	20.85	20.92	-0.07
Rhineland-Palatinate	4.92	0.42	4.50
Saarland	1.22	0.11	1.11
Saxony	5.13	4.69	0.44
Saxony-Anhalt	2.76	0.22	2.54
Schleswig-Holstein	3.78	0.29	3.49
Thuringia	2.59	0.22	2.37
Occupational sector			
S1: Production of goods	17.50	24.87	-7.37
S2: Personal services	20.34	25.03	-4.69
S3: Business admin./business services	42.27	31.76	10.51
S4: IT-sector and the natural sciences	10.35	5.14	5.21
S5: Others in commercial services	10.35	12.73	-2.38
Skill level			
1: Unskilled tasks	4.00	15.79	-11.79
2: Skilled tasks	36.38	51.91	-15.53
3: Complex tasks	26.15	15.61	10.54
4: Highly complex tasks	30.17	16.24	13.93
Working time			
Fulltime	79.44	70.10	9.34
Parttime	20.56	29.90	-9.34

This table shows participant characteristics of our final sample (N = 3,701) and the employment statistics from German Federal Employment Agency (BA 2024). Column (2) restricts the administrative data to employees aged 25–55, consistent with the age range of our sample. Our sample was drawn to match official population statistics concerning age, gender, and federal state. Occupational sector and skill level are classifications derived from KldB 2010. We define parttime if weekly working hours are <35.

Table 4.9: Beliefs about (non-)pecuniary returns

	All			Higher-skilled			Lower-skilled			Diff-in-Diff (10) DiD
	(1) With	(2) Without	(3) Diff	(4) With	(5) Without	(6) Diff	(7) With	(8) Without	(9) Diff	
Pecuniary returns: Individual expected earnings										
Absolute	5481.14 (2986.67)	5140.47 (2898.10)	-340.67*** (-4.98)	6542.64 (3352.78)	6150.73 (3297.32)	-391.90*** (-3.44)	4571.45 (2267.53)	4274.67 (2158.39)	-296.77*** (-4.23)	-95.13** (-3.28)
Non-pecuniary returns: General view										
Professional success	59.94 (32.12)	52.48 (32.23)	-7.46*** (-9.98)	63.25 (30.86)	53.61 (31.59)	-9.65*** (-9.03)	57.10 (32.90)	51.51 (32.74)	-5.59*** (-5.37)	-4.06*** (-4.93)
Non-pecuniary returns: Job characteristics										
Complex tasks	51.13 (32.79)	37.42 (29.25)	-13.70*** (-18.97)	54.36 (32.12)	39.81 (29.29)	-14.55*** (-13.83)	48.36 (33.11)	35.38 (29.06)	-12.98*** (-13.15)	-1.57 (-1.61)
Enjoy tasks	55.05 (30.39)	50.69 (29.89)	-4.36*** (-6.22)	57.15 (29.38)	49.39 (28.91)	-7.76*** (-7.78)	53.25 (31.12)	51.81 (30.67)	-1.44 (-1.47)	-6.32*** (-7.27)
Work-related stress	45.59 (31.03)	43.93 (31.71)	-1.66* (-2.28)	46.95 (30.13)	46.16 (31.28)	-0.80 (-0.76)	44.43 (31.75)	42.03 (31.96)	-2.40* (-2.38)	1.61 (1.81)
Non-pecuniary returns: Career outcomes										
Get promoted	36.83 (35.16)	26.50 (30.18)	-10.33*** (-13.56)	42.85 (35.81)	31.25 (31.42)	-11.60*** (-10.06)	31.67 (33.76)	22.43 (28.46)	-9.24*** (-9.34)	-2.35** (-2.79)
Quit job	27.28 (30.44)	30.69 (32.25)	3.41*** (4.68)	28.76 (30.61)	33.08 (32.64)	4.32*** (3.99)	26.02 (30.25)	28.65 (31.78)	2.63** (2.68)	1.69* (2.35)
Be laid off	15.24 (22.41)	18.50 (25.38)	3.27*** (5.87)	14.82 (21.68)	18.31 (24.88)	3.49*** (4.37)	15.59 (23.02)	18.67 (25.80)	3.08*** (3.97)	0.41 (0.67)

Notes: This table reports mean beliefs for the full sample and separately by skill level. Columns 1–3 show results for all respondents, Columns 4–6 for higher-skilled, and Columns 7–9 for lower-skilled individuals. Within each group, the first two columns report mean beliefs under the scenarios with and without firm training. The third column gives the mean difference between the two, with a paired *t*-test for statistical significance (p-value shown below the mean difference). Column 10 reports the difference-in-differences between higher- and lower-skilled respondents, based on an independent two-sample *t*-test. Beliefs are reported on a 0–100 scale, except for expected earnings (in euros). Significance levels: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

Table 4.10: Robustness: difference in return beliefs by workers' skill-level

Higher-skilled coefficient with controls for...	Pecuniary returns		Non-pecuniary returns						
	(1) Earnings	(2) Relative earnings	(3) Professional success	(4) Complex tasks	(5) Enjoy tasks	(6) Work-related stress	(7) Get promoted	(8) Quit job	(9) Be laid off
... none	95.13*** (28.96)	0.00 (0.01)	4.06*** (0.82)	1.57 (0.98)	6.32*** (0.87)	-1.61* (0.89)	2.35*** (0.84)	-1.69** (0.72)	-0.41 (0.62)
... demographics & preferences	-13.54 (38.80)	-0.03* (0.02)	2.21** (0.88)	1.28 (1.04)	4.11*** (0.91)	-2.02** (0.96)	0.98 (0.93)	-1.07 (0.80)	0.14 (0.69)
+ firm characteristics & sector FE	-24.24 (38.85)	-0.03** (0.02)	1.84** (0.89)	1.39 (1.06)	3.79*** (0.93)	-1.49 (0.97)	1.19 (0.93)	-1.12 (0.80)	0.25 (0.69)
+ training characteristics	-34.96 (39.25)	-0.03** (0.02)	0.92 (0.88)	0.65 (1.05)	2.29** (0.92)	-1.13 (0.99)	0.47 (0.93)	-0.57 (0.81)	0.58 (0.70)

Notes: Each column shows the estimated OLS coefficient on the indicator *higher-skilled*, with the corresponding belief gap as the dependent variable. Columns 1 and 2 refer to belief gaps in pecuniary returns; columns 3 to 7 refer to non-pecuniary returns. Average belief gaps are defined as ... Each row corresponds to one of four increasingly rich control sets. ... none: no additional covariates. ... demographics & preferences: adds age, gender, ..., and preference measures (risk, time, competitiveness). + firm characteristics & sector FE: further includes firm size and five dummies for the sectors (KIDB). + training characteristics: additionally controls for [insert, e.g., training duration, certification, or off-the-job indicators]. Robust standard errors are reported in parentheses. Significance levels: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

Table 4.11: Respondent characteristics separately for higher- and lower-skilled

	(1) All	(2) Higher-skilled	(3) Lower-skilled	(4) Difference
Age	40.78	39.28	42.08	2.80***
Female	0.50	0.43	0.56	0.13***
Weekly hours	36.35	37.08	35.73	-1.35***
Weekly hours <35	0.21	0.17	0.24	0.07***
Higher-skilled	0.46	1.00	0.00	-1.00
Professional Degree				
Vocational training	0.35	0.00	0.65	0.65***
Technical School	0.16	0.18	0.14	-0.04***
University	0.45	0.80	0.14	-0.66***
None/other	0.04	0.01	0.06	0.05***
Number of workers at firm	452.95	479.95	429.81	-50.14***
Years at firm	10.78	9.99	11.45	1.45***
Gross monthly earnings	4669.84	5618.94	3862.75	-1756.20***
Observations	3701	1708	1993	3701

Notes: N = 3701. This table displays averages for the whole sample and by skill-level of the respondents. Column 4 gives the mean difference between the two groups and stars denote significance from t-tests comparing group means. Gross monthly earnings are converted to full-time equivalent earnings and winsorized from below at the minimum wage (2,135 €) and from above at the 98th percentile of the German wage distribution (14,583 €).

Table 4.12: Firm training participation separately for higher- and lower-skilled

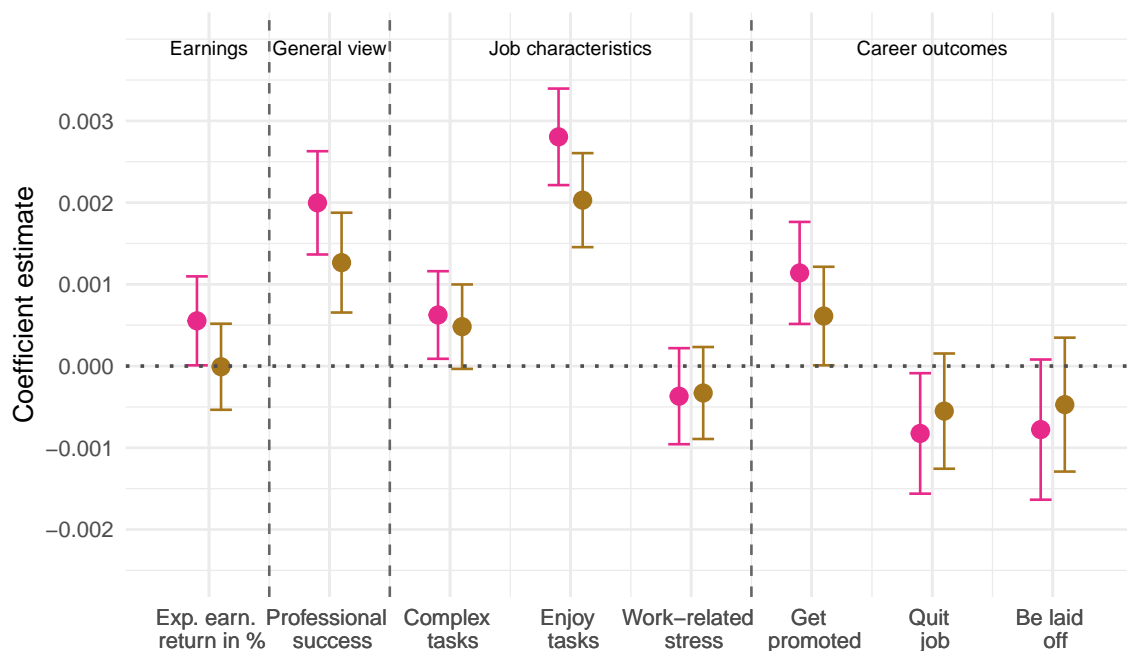
	(1) All	(2) Higher-skilled	(3) Lower-skilled	(4) Difference
Participated in the last 12 months	55.53	67.15	45.56	-21.60***
Conditional on participation				
Hours in training	37.92	45.19	28.74	-16.46*
Training financed by employer	0.98	0.98	0.97	-0.01
Training during working hours	0.96	0.98	0.94	-0.03***
Probability to participate in the next 12 months	31.57	38.25	25.85	-12.40***
Observations	3701	1708	1993	3701

Notes: Column 1 shows means for the whole sample; columns 2–3 show means by skill level. Column 4 gives the difference between groups; stars denote significance from t-tests comparing group means. Significance: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

Table 4.13: Details on firm training participation separately for higher- and lower-skilled

	(1) All	(2) Higher-skilled	(3) Lower-skilled	(4) Difference
Whole sample				
Job requires more skills than before	0.65	0.72	0.60	-0.12***
Firm training ever denied?	0.25	0.30	0.21	-0.09***
Possibility to learn without formal firm training	0.58	0.61	0.56	-0.05**
Number of available skill trainings (0-4)	1.64	1.88	1.43	-0.46***
No skill training available	0.22	0.16	0.28	0.12***
Firm training participants				
(Jointly) initiated firm training	0.66	0.71	0.59	-0.12***
Productivity increase through firm training (in %)	11.86	12.75	10.88	-1.87***
Learned skills are transferable	0.84	0.87	0.80	-0.07***
Skills learned at firm training:				
Soft skills	0.45	0.51	0.39	-0.12***
IT skills	0.40	0.47	0.32	-0.15***
Admin skills	0.25	0.22	0.29	0.06***
Specific skills	0.45	0.44	0.46	0.03
Number of skills learned (0-4)	1.62	1.70	1.54	-0.16***
Observations	3701	1708	1993	3701

Notes: This table provides means for the whole sample and by skill level of the respondents. Column 4 gives the mean difference between the two groups and stars denote significance from t-tests comparing group means. Significance levels: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

Figure 4.5: Do return beliefs predict realized firm training participation?

Controls ● None ● Worker-, firm-, and training characteristics, and sector fixed effects

Notes: Coefficient estimates from a regression of binary firm training participation in the last 12 months on the expected earnings return in % in the leftmost column and on the difference in values between the two hypothetical scenarios in percentage points for the remaining non-pecuniary belief dimensions. "None" indicates point estimates from a bivariate regression, while "Worker, firm, and detailed training characteristics + sector FE" indicates point estimates upon including these controls. Worker and firm characteristics: workers' age, gender, risk aversion, patience, locus of control, and current full-time equivalent earnings, as well as dummies indicating part-time work and above median firm size. Detailed training characteristics: indicators for the absence of training opportunities at the current workplace, whether workers were denied training previously, whether workers think their job is more demanding than a few years ago, and whether it is possible to accumulate skills through learning-by-doing. Fixed effects for five sectors are included.

Figure 4.6: Difference in return beliefs by workers' available skill training



Notes: This figure shows average belief gaps between individuals who have access to different types of skill training for each event. Respondents were asked: "For which types of skills does your employer offer you the opportunity to learn or improve them through firm training?" Answer options included: (1) soft skills and personal development, (2) IT skills, (3) office and administrative skills, (4) technical, practical, or job-specific skills, (5) none, and (6) other. Bars show group means with 95% confidence intervals. The number above each pair of bars indicates the mean difference between groups and stars denote significance levels from independent t-tests. Significance levels: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

Table 4.14: Skill gap in realized firm training

Dep. Var.: firm training in past year ($\times 100$)	(1)	(2)	(3)	(4)
Higher-skilled	21.595*** (1.600)	19.804*** (1.600)	9.937*** (1.586)	9.423*** (1.590)
Beliefs		✓		✓
Worker and firm char.			✓	✓
Training char.			✓	✓
Sector			✓	✓
R^2	0.047	0.066	0.248	0.253

Notes: $N = 3,701$. OLS regressions of a dummy for firm-training participation in the past year (multiplied by 100) on a higher-skilled indicator and varying controls. *Beliefs*: belief gaps in pecuniary and non-pecuniary dimensions. *Worker and firm char.*: age, gender, risk aversion, patience, locus of control, current full-time-equivalent earnings, dummies for part-time work and above-median firm size. *Training char.*: indicators for absence of training opportunities at the current workplace, past denial of training, perceived increase in job demands, and learn-by-doing possibilities. *Sector*: fixed effects for five sectors. Robust standard errors in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

4.B Appendix B: Information Experiment

Table 4.15: Balancing

	All (1)	Treatment (2)	Control (3)	p-value (4)
Age	40.78	40.71	40.85	0.617
Female	0.50	0.48	0.51	0.050
Weekly hours	36.35	36.39	36.31	0.689
Weekly hours <35	0.21	0.20	0.21	0.344
Higher-skilled	0.46	0.46	0.47	0.659
Professional Degree				
Vocational training	0.35	0.36	0.35	0.489
Technical School	0.16	0.15	0.17	0.058
University	0.45	0.45	0.44	0.481
None/other	0.04	0.04	0.04	0.939
Number of workers at firm	452.95	459.06	446.91	0.369
Years at firm	10.78	10.79	10.77	0.940
Gross monthly earnings	4669.84	4680.99	4658.79	0.793
Observations	3701	1839	1862	3701

Notes: Columns 1 to 3 show sample means for the indicated subgroup. Column 4 shows p-values from t-tests comparing the mean of each variable between treatment and control group. A joint F-test based on regressing a dummy that takes on value one for respondents in the treatment group on all covariates gives a p-value of 0.545.

Table 4.16: Prior beliefs

	All (1)	Higher-skilled (2)	Lower-skilled (3)	p-value (4)
Beliefs about (non-)pecuniary returns to firm training				
Firm training increases earnings	0.51	0.55	0.48	0.000
Firm training enhances career opportunities	0.74	0.76	0.72	0.002
Beliefs about learning				
Improvement in professional tasks through practice	0.80	0.82	0.78	0.001
Skills brought to the job are more important than effort	0.58	0.57	0.59	0.182
Ease of learning new things in firm training	0.73	0.79	0.68	0.000
Hesitation to attend training due to lack of recent learning	0.22	0.21	0.23	0.246
Indices				
Index: Positive return beliefs	-0.00	0.09	-0.08	0.000
Index: Positive learning beliefs	-0.00	0.14	-0.12	0.000
Observations	3701	1708	1993	3701

Notes: Columns 1 to 3 show sample means of the agreement to our prior belief questions for the indicated subgroup. Respondents were asked about their agreement to the statements on a 5-point likert scale. If a respondent (partially) agrees, the variable is coded as 1 and 0 otherwise. Column 4 shows p-values from t-tests comparing the mean of each variable between higher- and lower-skilled.

Table 4.17: Heterogenous treatment effects on training intentions by prior beliefs

	Probability to accept a training offer					
	.. with no costs			.. with costs		
	(1)	(2)	(3)	(4)	(5)	(6)
Treatment	2.234** (0.896)	2.791* (1.429)	3.134** (1.302)	2.138** (1.017)	1.944 (1.352)	3.104** (1.246)
Treatment x Positive return beliefs		-1.278 (1.777)			0.156 (1.999)	
Positive return beliefs		8.949*** (1.289)			12.191*** (1.439)	
Treatment x Positive learning beliefs			-2.731* (1.625)			-2.569 (2.099)
Positive learning beliefs			17.421*** (1.184)			8.445*** (1.477)
Treatment effect for positive (return/learning) beliefs		1.513 (1.050)	0.402 (0.968)		2.099 (1.472)	0.536 (1.686)
Control mean	81.03	81.03	81.03	30.70	30.70	30.70
Observations	3701	3701	3701	3701	3701	3701
Controls	✓	✓	✓	✓	✓	✓

Notes: OLS regressions. Dependent variable: *Probability of accepting a training offer of 80 hours of the employer with no costs: the employer fully covers the costs, or with costs: 20% of their current net monthly salary must be paid by themselves.* A standardized summary index is created from the first two items in Table 4.16 to capture prior beliefs about returns, while a second index summarizes the next four items to capture beliefs about learning. Both indices are standardized such that higher values indicate more positive beliefs regarding firm training; consequently, items 4 and 6 are reverse-coded. We then perform a median split on each index, creating dummy variables for "positive return beliefs" and "positive learning beliefs," which identify individuals with above-median, or more positive, beliefs in each dimension. Robust standard errors (clustered at the individual level) in parentheses. Significance of treatment effect for higher-skilled respondents are calculated using post-estimation Wald tests. Control mean: mean of the outcome variable in the control group. Controls include age, dummies for female and ten occupational major groups (KldB 2010), risk preferences and patience. Significance levels: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

Table 4.18: Heterogenous treatment effects by workers' age

	Probability to accept a training offer			
	.. with no costs		.. with costs	
	(1)	(2)	(3)	(4)
Treatment	2.234** (0.896)	1.822 (1.134)	2.138** (1.017)	2.821** (1.377)
Treatment x Age > median		0.881 (1.813)		-1.456 (2.040)
Age > median		-2.645 (1.939)		3.630 (2.234)
Treatment effect for age > median		2.702* (1.412)		1.365 (1.506)
Control mean	81.03	81.03	30.70	30.70
Observations	3701	3701	3701	3701
Controls	✓	✓	✓	✓

Notes: OLS regressions. Dependent variable: *Probability of accepting a training offer* of 80 hours of the employer *with no costs*: the employer fully covers the costs, or *with costs*: 20% of their current net monthly salary must be paid by themselves. Age > median is a dummy equal to 1 if the respondent is older than the sample median age (42). Robust standard errors (clustered at the individual level) in parentheses. Significance of treatment effect for higher-skilled respondents are calculated using post-estimation Wald tests. Control mean: mean of the outcome variable in the control group. Controls include age, dummies for female and ten occupational major groups (KldB 2010), risk preferences and patience. Significance levels: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

Table 4.19: Heterogenous treatment effects by workers' age (continuous)

	Probability to accept a training offer			
	.. with no costs		.. with costs	
	(1)	(2)	(3)	(4)
Treatment	2.234** (0.896)	-2.023 (4.437)	2.138** (1.017)	12.490** (5.124)
Treatment x Age		0.104 (0.107)		-0.248** (0.123)
Control mean	81.03	81.03	30.70	30.70
Observations	3701	3701	3701	3701
Controls	✓	✓	✓	✓

Notes: OLS regressions. Dependent variable: *Probability of accepting a training offer* of 80 hours of the employer *with no costs*: the employer fully covers the costs, or *with costs*: 20% of their current net monthly salary must be paid by themselves. Robust standard errors (clustered at the individual level) in parentheses. Significance of treatment effect for higher-skilled respondents are calculated using post-estimation Wald tests. Control mean: mean of the outcome variable in the control group. Controls include age, dummies for female and ten occupational major groups (KldB 2010), risk preferences and patience. Significance levels: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

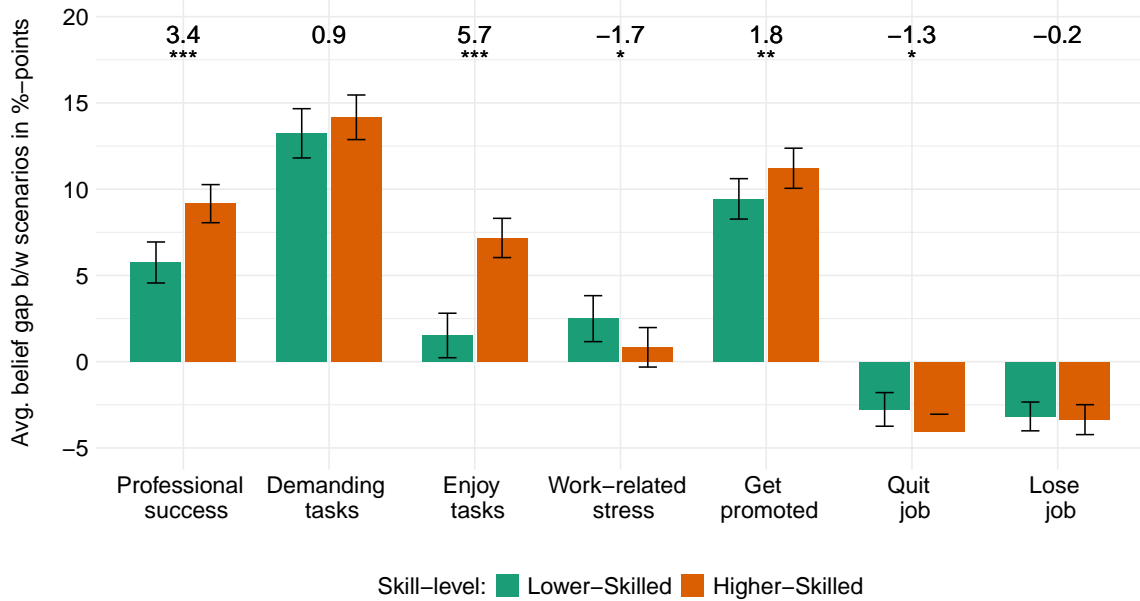
4.C Appendix C: Robustness to Alternative Definition of Higher-skilled

Table 4.20: Earnings expectations by workers' skill-level (alternative definition)

	Higher-skilled		Lower-skilled		Difference
	Mean (1)	Std. Dev. (2)	Mean (3)	Std. Dev. (4)	Lower – Higher (5)
Expected earnings in €					
With firm training	6,441	3,333	4,515	2,205	-1,926***
Without firm training	6,057	3,267	4,217	2,100	-1,840***
Individual expected returns					
Absolute	384	927	297	800	-86***
Relative in %	8.60	29.71	8.65	29.15	0.05
Share positive	0.49	0.50	0.53	0.50	0.04**
Observations		1,844		1,857	3,701

Notes: Mean and Std. Dev. display averages and standard deviations of worker-reported expected monthly gross earnings by workers' skill level (alternative definition). Individual expected returns average perceived individual absolute and relative returns across workers. Relative returns correspond to absolute returns divided by earnings in the scenario without firm training. "Lower – Higher" is the mean difference. Earnings expectations are converted to full-time equivalents, winsorized from below at the minimum wage (2,135€) and from above at the 98th percentile of the German wage distribution (14,583€), and then converted to full-time equivalent earnings. Significance levels: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

Figure 4.7: Difference in return beliefs by workers' skill-level (alternative definition)



Notes: This figure displays average belief gaps between lower- and higher-skilled (alternative definition) individuals for each event. Bars show group means with 95% confidence intervals. The number above each pair of bars indicates the mean difference between groups and stars denote significance levels from independent t-tests. Significance levels: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$

Table 4.21: Skill gap in firm training intentions (alternative definition)

Dep. Var.: Training Intentions	(1)	(2)	(3)	(4)
Higher-skilled	11.728*** (1.055)	10.430*** (1.039)	2.874*** (1.056)	2.659** (1.049)
Beliefs		✓		✓
Worker and firm char.			✓	✓
Training char.			✓	✓
Sector			✓	✓
R^2	0.032	0.077	0.213	0.232

Notes: $N = 3,701$. Coefficient estimates of a dummy indicating whether a worker is higher skilled (alternative definition) in a regression of the probability of participating in firm training next year on varying sets of controls. Beliefs: belief gaps in pecuniary and non-pecuniary dimensions. Worker and firm char.: workers' age, gender, risk aversion, patience, locus of control, and current full-time equivalent earnings, as well as dummies indicating part-time work and above-median firm size. Training char.: indicators for the absence of training opportunities at the current workplace, whether workers were denied training previously, whether workers think their job is more demanding than a few years ago, and whether it is possible to accumulate skills through learning-by-doing. Sector: fixed effects for five sectors. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

Table 4.22: Information treatment effects on training intentions (alternative definition)

	Probability to accept a training offer			
	.. with no costs		.. with costs	
	(1)	(2)	(3)	(4)
Treatment	2.234** (0.896)	2.480** (1.243)	2.138** (1.017)	1.782 (1.311)
Treatment x Higher-skilled		-0.516 (1.659)		0.828 (1.901)
Higher-skilled		2.564** (1.221)		4.880*** (1.381)
Treatment effect on higher-skilled		1.964* (1.189)		2.609* (1.474)
Control mean	81.03	81.03	30.70	30.70
Observations	3701	3701	3701	3701
Controls	✓	✓	✓	✓

Notes: OLS regressions. Dependent variable: *Probability of accepting a training offer of 80 hours of the employer with no costs: the employer fully covers the costs, or with costs: 20% of their current net monthly salary must be paid by themselves.* Robust standard errors (clustered at the individual level) in parentheses. Significance of treatment effect for higher-skilled respondents are calculated using post-estimation Wald tests. Control mean: mean of the outcome variable in the control group. Controls include age, dummies for female and ten occupational major groups (KldB 2010), risk preferences and patience. Significance levels: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

4.D Appendix D: Robustness to Winsorization

Table 4.23: Earnings expectations by workers' skill-level (no winsorization)

	Higher-skilled		Lower-skilled		Difference Lower – Higher
	Mean	Std. Dev.	Mean	Std. Dev.	
Expected earnings in €					
With firm training	10,689	22,933	8,117	113,572	-2,572
Without firm training	9,755	23,195	31,196	1,178,858	21,440
Individual expected returns					
Absolute	933	13,754	-23,079	1,184,367	-24,012
Relative in %	19.22	162	5,256	226,289	5,237
Share positive	0.55	0.50	0.57	0.50	0.01
Observations		1,993		1,708	3,701

Notes: Mean and Std. Dev. display averages and standard deviations of worker-reported expected monthly gross earnings by workers' skill-level (no winsorization). Individual expected returns average perceived individual absolute and relative returns across workers. Relative returns correspond to absolute returns divided by earnings in the scenario without firm training. "Lower – Higher" is the mean difference. Earnings expectations are converted to full-time equivalents. Significance levels: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

Table 4.24: Earnings expectations descriptives (no winsorization)

	Mean (1)	Std. Dev. (2)	Median (3)
Expected earnings in €			
With firm training	9304	84786	4686
Without firm training	21301	865188	4375
Individual expected returns			
Absolute	-11997	869153	171
Relative in %	2836	165957	4.35
Share positive	0.56	0.50	1

Notes: $N = 3701$. Means and standard deviations refer to worker-reported expectations of monthly gross earnings. *Individual expected returns* contain summary statistics of workers' perceived absolute and relative returns. Relative returns correspond to the within-worker difference in expected earnings across scenarios divided by earnings in the scenario without firm training. All figures are in full-time equivalents.

4.E Appendix E: Robustness to Bad Data

In this section we report our main results only for those individuals who are a reliable type according to the definition by Dohmen and Jagelka 2024.

Table 4.25: Earnings expectations descriptives (only reliable respondents, $N = 2962$)

	Mean (1)	Std. Dev. (2)	Median (3)
Expected earnings in €			
With firm training	5497	2888	4800
Without firm training	5156	2798	4461
Individual expected returns			
Absolute	340	781	100
Relative in %	8.32	25.76	2.27
Share positive	0.51	0.50	1

Notes: $N = 3701$. Means and standard deviations refer to worker-reported expectations of monthly gross earnings. *Individual expected returns* contain summary statistics of workers' perceived absolute and relative returns. Relative returns correspond to the within-worker difference in expected earnings across scenarios divided by earnings in the scenario without firm training. All figures are in full-time equivalents. Earnings expectations are winsorized from below at the minimum wage (2,135€) and from above at the 98th percentile of the German wage distribution (14,583€).

Table 4.26: Earnings expectations by workers' skill-level (only reliable respondents, $N = 2962$)

	Higher-skilled		Lower-skilled		Difference
	Mean (1)	Std. Dev. (2)	Mean (3)	Std. Dev. (4)	Lower – Higher (5)
Expected earnings in €					
With firm training	6,571	3,206	4,609	2,236	-1,962***
Without firm training	6,187	3,153	4,305	2,118	-1,882***
Individual expected returns					
Absolute	384	821	304	745	-80***
Relative in %	8.08	23.73	8.53	27.33	0.44
Share positive	0.49	0.50	0.53	0.50	0.04**
Observations		1993		1708	3701

Notes: Mean and Std. Dev. display averages and standard deviations of worker-reported expected monthly gross earnings by workers' skill-level. Individual expected returns average perceived individual absolute and relative returns across workers. Relative returns correspond to absolute returns divided by earnings in scenario without firm training. "Lower – Higher" is the mean difference. Earnings expectations are converted to full-time equivalents as well as winsorized from below at the minimum wage (2,135€) and from above at the 98th percentile of the German wage distribution (14,583€). Significance levels: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

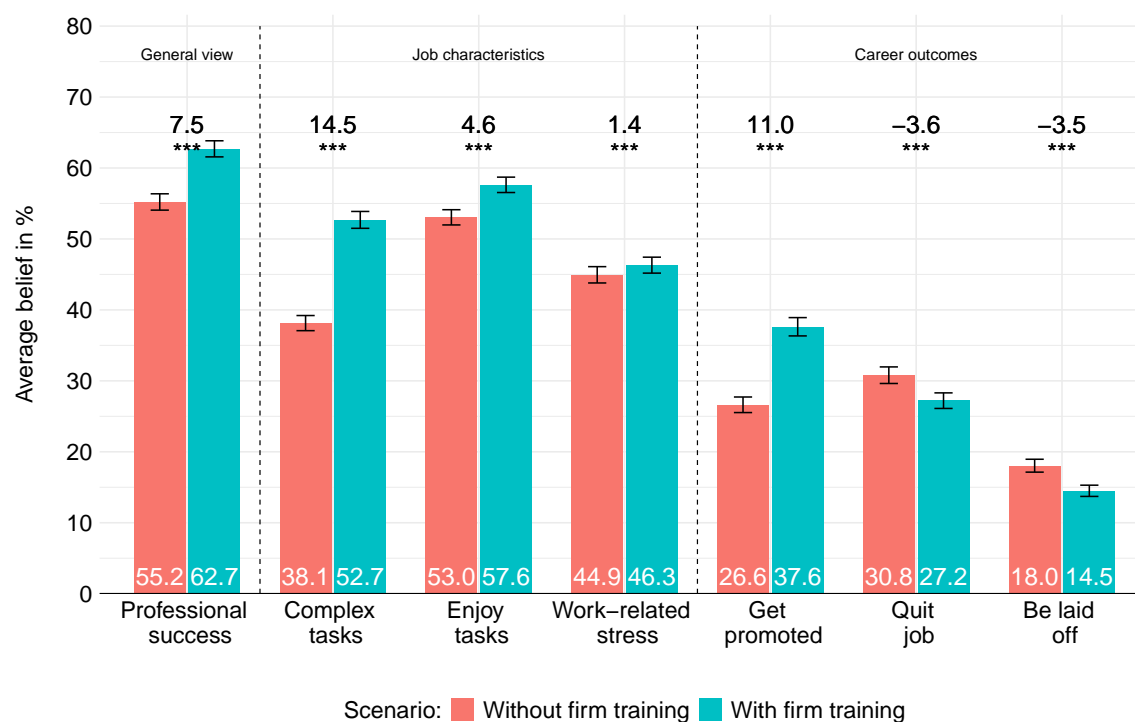
Figure 4.8: Beliefs about non-pecuniary returns (only reliable respondents, $N = 2962$)

Figure 4.9: Difference in return beliefs by workers' skill-level (only reliable respondents, N = 2962)

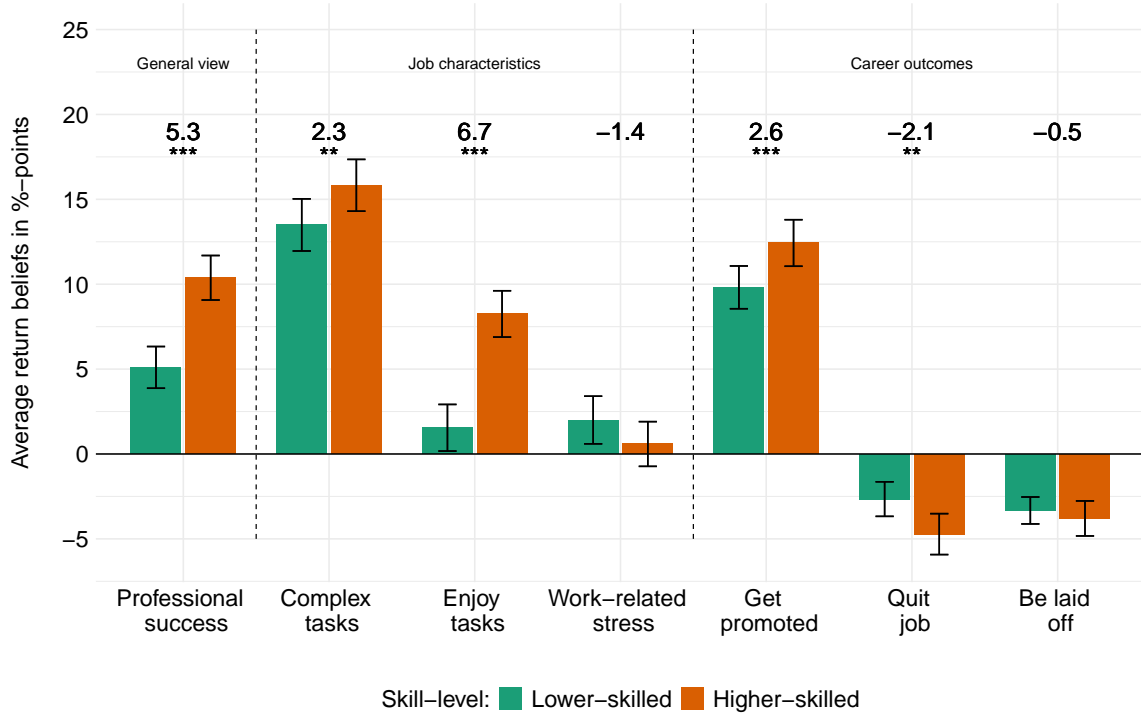


Table 4.27: Information treatment effects on training intentions (reliable respondents, $N = 2962$)

	Probability to accept a training offer			
	.. with no costs		.. with costs	
	(1)	(2)	(3)	(4)
Treatment	1.695*	2.875**	1.977*	3.216**
	(0.991)	(1.441)	(1.167)	(1.541)
Treatment x Higher-skilled		-2.499		-2.524
		(1.982)		(2.343)
Higher-skilled		4.317**		7.138**
		(1.438)		(1.714)
Treatment effect on higher-skilled		0.376		0.692
		(1.343)		(1.767)
Control mean	82.75	82.75	30.50	30.50
Observations	2962	2962	2962	2962
Controls	✓	✓	✓	✓

Notes: OLS regressions. Dependent variable: *Probability of accepting a training offer* of 80 hours of the employer *with no costs* (employer fully covers costs) or *with costs* (20% of current net monthly salary is self-paid). Robust SEs (clustered at the individual level) in parentheses. Treatment effect for higher-skilled from post-estimation Wald tests. Control mean: mean outcome in control group. Controls: age, female, ten occupational major groups (KldB 2010), risk preferences, patience. Significance: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

4.F Appendix F: Survey

This section gives an overview over the main questions of our survey. Our measurement of firm training participation, the belief elicitation and the information experiment.

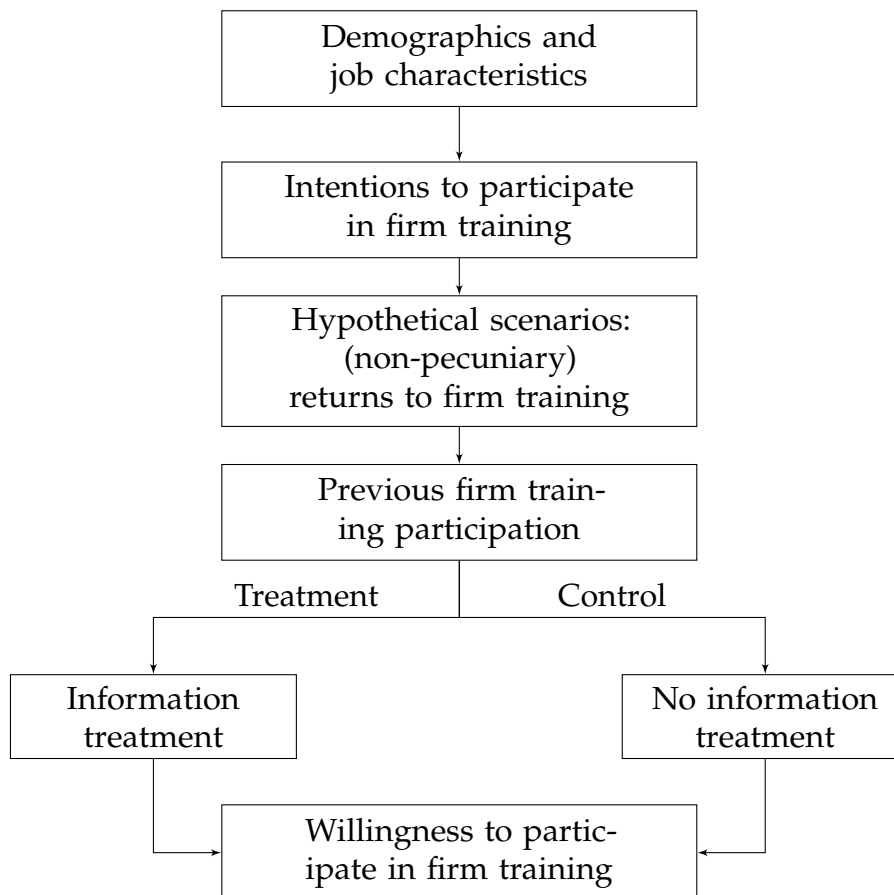


Figure 4.10: Survey Flow

4.F.1 Measuring Firm Training Participation

[firm training participation:]

In the following, we ask you a few questions about firm training.

Firm training are courses or events offered by your company to refresh existing professional skills or learn new skills. They are fully or partially financed by your company and can take place both during and outside working hours. Firm training can be provided both externally and internally and can last from a few hours to several months.

How many firm training courses have you attended in the last 12 months?

Please do not include informal learning from your colleagues or learning through experience. Please enter "0" if you have not taken part in any firm training.

- firm trainings

[new page:]

Now, we would like to ask about the time you have spent on firm training in the past 12 months. How many full hours have you spent on firm training in total over the last 12 months?

If you do not know exactly, please estimate.

[new page:]

How many days were the <embedded field> training courses that you took part in last year, amounting to <embedded field> hours, divided into?

[new page:]

Have you already taken part in firm training since you started working for your current employer?

- Yes/No

[new page, firm training intention:]

In the next questions, please rate how likely it is that certain events will occur. To do this, we ask you to enter a number between 0% and 100%. The percentage values correspond to the probability that an event will occur. 0% means completely ruled out, and 100% means absolutely certain. You can choose any number between 0% and 100% that you think best corresponds to the probability of the event.

How likely is it that you ...

- ... participate in firm training totaling at least 40 hours (equivalent to about 1 week full-time) in the next year? ... % (0 - 100)
- ... participate firm training of at least 40 hours per year (equivalent to about 1 week full-time) every year until the end of your working life? ... % (0 - 100)
dont show because we dont use it?

4.F.2 Belief Elicitation

[non-pecuniary beliefs:]

We now ask you to think about the future of your professional life. This question is of great importance to us, so please answer it as accurately as possible.

Please imagine the following two scenarios:

<p style="text-align: center;">Scenario 1</p> <p>Every year until the end of your working life, you take part in firm training totaling 80 hours per year (equivalent to around 2 weeks full-time).</p>	<p style="text-align: center;">Scenario 1</p> <p>You do not take part in any firm training until the end of your working life.</p>
--	---

Please answer the following questions for the two scenarios. We ask you to enter a number between 0% and 100%. 0% means completely ruled out, and 100% means absolutely certain. You can use the values in between to scale your answer.

	With firm training (0 - 100)	Without firm training (0-100)
Now, we would like to ask about your expectations for the year 2030. How likely is it that you...		
... enjoy your day-to-day professional tasks in 2030?		
... have a lot of stress at work in 2030?		
... take on more demanding tasks in 2030 than today?		
... be promoted at least once by the year 2030?		
... quit your current job by the year 2030?		
... be laid off from your current job by 2030?		
... be successful in your professional life in 2030?		

[new page, earnings beliefs:]

Now, we would like to ask about your expectations regarding your salary development.

Please continue to imagine the two scenarios.

Assume that you will work the same number of hours as you do today. Please disregard the effects of inflation on wages. This means that one euro today is worth

the same as one euro in the year 2030. What do you think your average gross monthly salary will be in 2030 in both scenarios?

Your current gross monthly salary: [insert value]

	With firm training (0 - 100)	Without firm training (0-100)
Expected gross monthly salary in 2030		

4.F.3 Information Experiment

[prior beliefs:]

In the following, we would like to ask you a few questions about learning in the workplace and continuing professional development.

How much do you agree with the following statements? (I strongly agree, I tend to agree, neither, I tend to disagree, I strongly disagree)

- By participating in firm training, my salary increases in the long term.
- Taking part in firm training improves my career opportunities in the long term.
- I can get better at my professional tasks by practicing.
- More important than all efforts are the abilities one possesses.
- It would be easy for me to learn new things as part of further training.
- The fact that I am no longer used to learning prevents me from taking part in firm training.

[new page, information treatment:]

Participation in further training ensures that employees' skills keep pace with technological change. Studies of the German labor market have shown that firm training protects against job loss and often leads to higher wages in the long term.

After a long break from learning, learning new skills through firm training can seem challenging. However, numerous studies show that you can improve or relearn your skills through regular practice well into old age.

[new page, probability of accepting training offer:]

Imagine your employer offers you the opportunity to take part in 80 hours of firm training (equivalent to about 2 weeks full-time) that takes place during working hours. Please enter a number between 0% and 100% for each of the following questions. 0% means completely ruled out, and 100% means absolutely certain. You can choose any number between 0% and 100% that you think most closely corresponds to the probability of the event.

How likely is it that you will accept the offer...

- ... if the employer fully covers the costs of the firm training? ...%
- ... if part of the costs of firm training amounting to 20% of your current net monthly salary have to be payed by yourself: ... %

5

Labor Market Returns to Skills: Evidence from LinkedIn

5.1 Introduction

Skills—and human capital more broadly—are primary determinants of earnings differences between workers and, thus, of the overall earnings distribution (Becker 1962; Mincer 1974). Workers invest in schooling or training, often forgoing earnings, to enhance their productivity and earn higher wages over their careers. While returns to both secondary and tertiary education are well documented (Card 1999; Heckman et al. 2006b; Kirkeboen et al. 2016; Chetty et al. 2023), much less is known about the returns to skills acquired on the job once formal schooling is completed. This gap is notable given the substantial contribution of post-school skills to the overall stock of human capital and their importance in explaining divergent earnings trajectories across workers, in particular early in the career (Rubinstein and Weiss 2006; Black et al. 2023; Jedwab et al. 2023). While a growing literature acknowledges that jobs differ in the learning opportunities they provide (e.g., Rosen 1972; Gibbons and Waldman 2006; Arellano-Bover and Saltiel 2024), direct evidence on on-the-job skill accumulation remains scarce, as most worker-level panel datasets lack information on how workers' skill sets evolve over time.¹As a result, even studies that

¹A rare exception is the panel dimension of the OECD's PIAAC skill data in Germany used in Arellano-Bover (2022) and Hanushek et al. (2025) which, however, lacks the detailed information on career trajectories.

emphasize the role of on-the-job learning in explaining earnings dynamics must infer it indirectly—without observing skills themselves (e.g., Adda and Dustmann 2023).

In this paper, I provide evidence on the career returns to accumulating skills on the job. I use longitudinal self-reported resumes including skill sets of more than five million college graduates in the US on LinkedIn. I proxy skill accumulation on the job using within-worker variation in reported skills on profiles between 2019 and 2021. I find sizable returns to skill accumulation on the attainment of highly-paid jobs. Workers adding skills have increased job mobility with regards to both positions and occupations alike. This increased mobility goes along with higher pay that is increasing over time. My findings suggest that skill acquisition may be an important pathway to upward mobility in the labor market.

I build a panel of college graduates who were employed in the US and maintained a LinkedIn profile continuously from 2015 to 2023. While workers' positions including occupation data and predicted earnings are available for the entire period, changes in their listed skill sets are only observable between 2019 and 2021. I measure workers' mobility using position and occupation transitions, and leverage two different earnings imputations that capture different parts of the overall variation in earnings. First, I use worker-level predicted earnings based on position characteristics such as job title, firm, location, and year. Second, I use occupation-level average earnings from the American Community Survey (ACS), which provides a benchmark tied to survey-reported wages and abstracts from within-occupation heterogeneity. While both earnings imputations have limitations when compared to earnings from administrative data, they are informative about which workers transition into more highly-paid positions and occupations. Since both earnings measures vary meaningfully only upon changes in position or occupation, it is important to study these outcomes simultaneously.

To identify the effect of on-the-job skill accumulation on career trajectories, I exploit the unusually detailed information on educational and career trajectories available on LinkedIn profiles. This detail allows me to compare workers who acquire at least one new skill ("treated") to "control" workers who do not, despite having highly comparable resumes and earnings prior to skill accumulation. My identification strategy relies on a selection-on-observables assumption (Rosenbaum and Rubin 1983), arguing that unobserved confounding is unlikely to be substantial given the exhaustiveness of pre-treatment information. Ideally, whether or not a worker accumulates skills during that period is as good as random conditional

on pre-treatment controls and outcomes. While not the only possible mechanism, this idea is consistent with the notion that some work environments are more conducive to learning than others, and assignment to such environments is random conditionally. Although workers' choice whether or not to add skills to their profiles is not random (not even conditionally), it might be unrelated to the counterfactual career outcomes of interest. I estimate propensity scores for sample trimming to ensure overlapping distributions in pre-treatment variables across treatment and control groups (Imbens and Rubin 2015).

I show that accumulating skills has sizable and growing effects on the attainment of more highly-paid jobs in two steps. First, I demonstrate that workers adding one or more skill(s) to their profiles are more likely to switch job positions and occupations, and that this mobility is geared towards employment with higher pay. Between 2019 and 2023, treated workers make 0.078 or 11.2% more switches to job positions than control workers. Switches to positions with above-average year-on-year changes in earnings are up by 0.049 or 12.1%. Turning to occupations, these numbers are 0.025 and 0.016, corresponding to 6.6% and 8.5% higher switching rates to another occupation and to a higher-paying occupation, respectively.

Second, I quantify earnings effects. Jobs held by workers accumulating skills have gross yearly predicted earnings premia of 650\$ in 2021 and 882\$ in 2023, corresponding to earnings increases of 0.74% and 0.94%, respectively, relative to comparable workers who did not add skills. In line with the lower mobility between occupations compared to job positions, effect sizes using ACS occupation level average earnings are smaller in magnitude but reveal the same pattern. In particular, earnings increases amount to 328\$ (0.34%) in 2021 and 635\$ in 2023 (0.61%).

To bolster the credibility of the estimated effects, I show that there is no sizable placebo effect on lagged earnings. Additionally, I conduct sensitivity analyses to assess the role of unobserved confounding. Placebo effect estimates are small in magnitude and merely borderline significant, bolstering the credibility of the unconfoundedness assumption. Sensitivity analyses suggest that the vast set of pre-treatment characteristics render it unlikely that omitted variables fully account for the estimated effects. In particular, a hypothetical confounder would need to have strong explanatory power for residual variation in reporting additional skills and career outcomes to overturn effect estimates.

In terms of robustness, I show that using different definitions of skill accumulation still yield plausible results. In particular, decomposing the total effect of

accumulating skills into contributions along the intensive and extensive margins demonstrates that the extensive margin effect dominates. Furthermore, estimates are insensitive to dropping parts of the sample with potentially confounded variation in skill reporting.

While the LinkedIn data provide unique advantages in that they contain detailed information on educational credentials, job histories, and skills, they also suffer from limitations. In particular, I only observe variation in workers' skill sets from 2019 through May 2021. In addition to being rather short, the period covers the onset of the Covid-19 pandemic, potentially affecting the external validity of my findings. Finally, skill depreciation (e.g., Cohen et al. 2025; Hanushek et al. 2025) is not observable in the data, constraining the variation I can exploit to investigate the relationship between skill accumulation and career outcomes.

This paper contributes to three strands of literature. First, I contribute to the literature on the sources of wage growth (Rubinstein and Weiss 2006; Adda and Dustmann 2023), in which human capital accumulation—especially on the job—plays a central role (e.g., Gathmann and Schönberg 2010). In particular, I exploit that some job environments are more conducive to on-the-job learning than others, and show that resulting skill accumulation has lasting effects on career trajectories and earnings. The idea that job environments differ in their on-the-job learning opportunities goes back to Rosen (1972), who investigated firms and has recently received renewed attention (Gregory 2020; Arellano-Bover 2024). I build on this line of research by directly observing changes in workers' skill sets—rather than inferring them from wage growth or task-based proxies—and by linking these changes to subsequent earnings gains (Arellano-Bover and Saltiel 2024; Adda and Dustmann 2023). I also show that skill accumulation increases job mobility, which itself has been identified as a key driver of wage growth over the life cycle (e.g., Topel and Ward 1992).

Second, I contribute to the literature on labor market returns to skills (summarized in Deming and Silliman 2024 and Woessmann 2025) by leveraging repeated observations of workers' skill sets to identify the effect of on-the-job skill accumulation on subsequent earnings trajectories. Much of the existing evidence relies on cross-sectional correlations or links earnings trajectories to static snapshots of skill portfolios, limiting causal interpretation. In contrast, I exploit within-worker changes in skills over time and relate them to subsequent labor market outcomes, conditional on rich pre-treatment covariates. This panel structure allows me to isolate the role

of skill accumulation and to demonstrate downstream career effects using multiple post-treatment years. My approach complements other efforts to estimate returns to skills using (quasi-)experimental variation, such as resume audits (Piopiunik et al. 2020) or institutional reforms (Hanushek et al. 2015), by relying on rich pre-treatment job histories to identify effects.² Dorn et al. (2025) leverage a different subset of the data used here. The paper documents the value of LinkedIn-based skill measures for capturing multidimensional human capital. Skills can explain larger shares of the variation in earnings than more traditional measures of human capital. Larger numbers of skills as well as higher shares of occupation-specific and managerial skills are associated with higher earnings. My work shows that within-worker across-time variation in skills affects earnings, thereby complementing the evidence using cross-sectional between-worker variation in that paper.

Finally, I contribute to the literature on on-the-job learning, which includes firm training involving forgone production as well as learning-by-doing (Black et al. 2023). In the latter, skill accumulation is a byproduct of employment and does therefore not necessitate costly investments. While I remain agnostic about the precise mechanism—whether skills are acquired through formal training or informal learning—the paper provides novel evidence on the returns to skill accumulation on the job. Unlike prior studies that infer learning from productivity and wage dynamics (e.g., Konings and Vanormelingen 2015), I observe changes in workers' skill sets directly, offering a rare opportunity to quantify the wage effects of newly accumulated skills. I also show that skill accumulation increases job mobility that goes along with higher pay, supporting the view that these skills are at least partially general in nature (e.g., Becker 1962; Acemoglu and Pischke 1998).

5.2 Data and Descriptive Statistics

This section details the sources of the data, rules for inclusion in the analysis sample, and provides descriptives for skills and earnings data. Since parts of the data have been used in previous work (Dorn et al. 2025), I focus on the departures from that setting. The description of the 2019 cross-section and the variables obtained

²While not the focus of this paper, many papers investigating labor market effects of government-sponsored active labor market programs catered to unemployed workers use this type of identification strategy (e.g., LaLonde 1986; Lechner et al. 2011; Biewen et al. 2014). For a summary of this literature, see McCall et al. (2016). In addition, Ruhose et al. (2019) use such an identification strategy to investigate the effect of work-related training.

therefrom largely follows that work, which provides additional detail, in particular on the skill clusters and their subsequent classification into skill classes.

Description of data other than skills and earnings, which are exclusively used for estimation of propensity scores (see Section 5.3.1), can be found in Appendix 5.A.1.

5.2.1 LinkedIn Profiles

The main data source for this project is a set of repeated snapshots of publicly accessible LinkedIn profiles of US college graduates. These data are provided by Revelio Labs, a workforce intelligence company that compiles scraped profiles in a database and enriches them with further information, relying on proprietary algorithms. LinkedIn profiles are structured like resumes (see Figure 5.B.1 in the Appendix for an example), listing employment histories, educational credentials including fields of study, and self-reported skills. With its more than 239 million users in the US (LinkedIn 2025), it is the most important online professional network and is widely used by workers and employers alike. For example, 69% of users report finding their LinkedIn network extremely or very useful for their careers. Among those actively seeking a new position, over 80% use the platform to identify job opportunities and research prospective employers (Evsyukova et al. 2025). Employers, on the other hand, use LinkedIn to advertise vacancies and receive as well as screen applications. Skills play a central role in this ecosystem, allowing workers to signal abilities and enabling employers to assess fit. LinkedIn's features—including a skill overlap score between a candidate's profile and a job posting—underscore the platform's role as a high-stakes environment for workers seeking to advance their careers.

5.2.2 Panel Creation and Sample

Although LinkedIn profiles are structured like resumes, the construction of a panel dataset requires different approaches for skills compared to job and education histories. For positions and education data, profiles have a built-in panel dimension: each snapshot reveals the current job or degree as well as a full history, including user-reported start and end dates (see Figure 5.B.1). As a result, even annual snapshots allow for a precise reconstruction of job transitions. The skill section, by contrast, has no temporal dimension and instead reflects only the set of skills a worker chooses to display at the time of scraping. That is, changes to skill sets

can only be recorded if a profile is scraped repeatedly. In my dataset, skills are timestamped based on when they appeared first in successive snapshots. Generally, each snapshot contains all profile information that is visible when LinkedIn is accessed without being logged in.

Profiles were initially scraped between January 2019 and September 2019. January 2019 is the initial scrape month for 77% of the sample. During each of the months February, March, May, June, July, and September, between 4 and 5% of profiles in my sample were added. As shown in Figure 5.1, I extract workers' information for 2019 in the month their profile was scraped initially. This information includes a vast set of pre-treatment information in addition to positions data as well as skill sets. Pre-treatment covariates include detailed information on degrees and work experience over workers' careers (see section 5.3.1 for the full list). To account for the heterogeneity in scrape month, I include fixed effects for the initial scrape month in regression analyses as well as estimation of propensity scores. Subsequent snapshots of all profiles occurred approximately every other month.

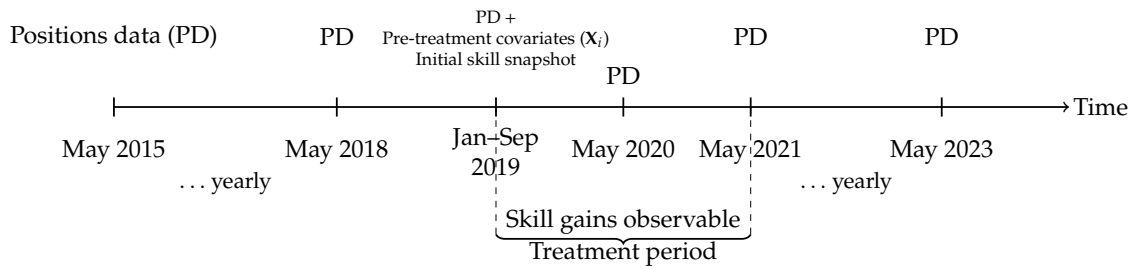
Even though skills are part of every snapshot, I only use workers' skill sets from two dates rather than leveraging the fine-grained timing of skill updates. These two dates are the initial scrape month in 2019 and May 31, 2021. May 31, 2021 is the date when skills disappeared from the publicly visible part of the profile. After this date, changes to skill sets are no longer visible. Finally, the structure of the database is such that only skill additions are visible.

In years other than 2019, I fix May 31 as the reference date for extracting job positions and education data. Anchoring annual snapshots to May 31 aligns snapshot dates with the day of the last date on which skill gains were observable (May 31, 2021). This alignment allows me to precisely date outcomes relative to the end of the treatment period—e.g., to measure what happens exactly one year before, during, or after this date. Using a consistent calendar date across years also ensures a uniformly spaced panel of job histories.

Positions data includes start date, predicted earnings, job title, firm, location, and occupation³ for the current position.

Regarding sample inclusion, I only keep workers who are employed in the US at that time and have non-missing gender, occupation, state, and education information, and are aged between 23 and 64. Furthermore, included workers

³Revelio Labs maps raw job titles to O*NET titles which are subsequently mapped to SOC 6 2018 occupation codes.

Figure 5.1: Timeline of sample construction

Notes: Positions data: Start date, job-based earnings, job title, firm, occupation. Time-invariant covariates (X_i): 5 highest degrees obtained, 39 fields of study, 5 college ranking brackets, years of job tenure, 22 occupations, 154 Metropolitan Statistical Areas, number of positions since 2015, 22 occupations at career start, graduation cohort, age in years, gender, ratio of actual to potential work experience, number of skills, skill shares for general, managerial, and specific.

have at least an Associate's degree, and non-missing skill data when they first appear in the database in 2019.

In addition, I condition on availability of position data in pre- and post-treatment years for sample inclusion. This means that profiles that do not report past positions or delete their profiles at some point by May 2023 are dropped from the sample. This is because pre-treatment position data is needed to estimate propensity scores, and conditioning on the availability of post-treatment outcomes for all observations rules out sample composition issues in the interpretation of effects.

Finally, I drop workers with estimated propensity scores outside [.05, .45] to ensure overlap (see section 5.3.4).

5.2.3 Skills

I use the 10,413 most commonly reported skills on LinkedIn profiles of workers in the US, which cover more than 90% of all skill entries. One feature that helps keeping this number rather low is that adding skills to profiles on LinkedIn is semi-structured: when users start typing, LinkedIn suggests different auto-completions based on popularity of the suggested skills among other users.

To put more structure on the skills, I conduct two dimension reduction exercises. First, as in Dorn et al. (2025), I make use of the word embedding algorithm to group similar skills into 48 skill clusters provided by Revelio Labs.⁴ Second, the 48 skill clusters are iteratively mapped to one of three skill classes. In particular, each skill

⁴Please refer to Dorn et al. (2025) for details.

clusters' Gini coefficient across fine-grained occupation categories is computed to classify skills as (occupation-)“specific” (upper tercile of Gini coefficients) or “general” (bottom two terciles of Gini coefficients). Finally, the 7 general skills that are most overrepresented among workers holding a top management position are labelled as “managerial” skills.

In the distinction put forward by Woessmann (2025), most skills on LinkedIn relate directly to performing a specific job task augmenting workers' marginal product in that task (“applied skills”) rather than representing basic skills such as cognitive skills. For instance, “Microsoft Office”, “customer service”, and “social media” are among the five most commonly reported skills. Some skills, however, are clearly non-cognitive in the sense that they govern what kinds of tasks workers choose to work on (Deming 2023b). In particular, the 15 most reported skills include “leadership”, “(people) management”, and “strategic planning”.

Table 5.1 details the variation in skills. In 2019, workers report 20 skills on average, roughly half of which are general. Specific and managerial skills each take a quarter. 17.6% of workers add at least one skill to their profiles between 2019 and May 2021 and therefore constitute the treatment group. Most workers add skills only once. Only 1.3% of the sample or 7.4% of the treatment group add skills twice or more often (the maximum is five). Overall, workers add half a skill on average over this time span, reflecting the fact that the number of skills added is zero-inflated. A majority of treated workers adds exactly one skill (62.3%), and more than three quarters add a single or two skills. Since there are some very large numbers of skills added, the average number of skills added is 3. Only 12.5% of workers add more than five skills.

Table 5.3 shows that treated workers, on average, list four more skills pre-treatment than control group workers. In addition, treated workers report a one (four) percentage point higher share of managerial (specific) skills pre-treatment. The average share of general skills reported is six percentage points lower in the treatment group.

5.2.4 Interpreting Skill Additions as Human Capital Accumulation

Changes in workers' self-reported skills on LinkedIn are the empirical counterpart of changes in a worker's latent multidimensional skill vector. This subsection

Table 5.1: Descriptive statistics for across-time variation in skills

	Mean	Std. Dev.	Min	Max	Median
# of skills 19	19.775	10.417	1	108	18
Δ # of skills	0.527	2.385	0	101	0
Adds skills	0.176	0.381	0	1	0
Adds skills ≥ 2 times	0.013	0.112	0	1	0
Share general	0.519	0.319	0	1	0.533
Share managerial	0.232	0.246	0	1	0.150
Share specific	0.249	0.310	0	1	0.091
<i>Conditional on adding skills:</i>					
Cond. Δ # of skills	2.994	4.995	1	101	1
# of times skills added	1.189	0.423	1	5	1
Adds 1 skill	0.623	0.485	0	1	1
Adds 2 skills	0.140	0.347	0	1	0
Adds 3 to 5 skills	0.112	0.315	0	1	0
Adds > 5 skills	0.125	0.331	0	1	0
Observations					5141520
Obs. add. skills					904301

Notes: Summary statistics for key skill variables. “# of skills 19” refers to the total number of skills listed on a worker’s profile in 2019. “ Δ # of skills” is the change in the number of skills between 2019 and 2021. “Adds skills” is an indicator for whether a worker added at least one skill over this period. “Adds skills ≥ 2 times” indicates whether skill additions occurred in at least two distinct snapshots. “Share general”, “Share managerial”, and “Share specific” denote the fraction of each skill type among all listed skills in 2019. The lower panel restricts to the subset of workers who added at least one skill. “Cond. Δ # of skills” reports the change in number of skills among this group. “# of times skills added” is the number of times a worker added at least one skill. The next four rows break down the number of skills added into mutually exclusive categories.

discusses the validity of this measure for classifying treatment and control groups and argues that skill additions can be given a human capital interpretation.

Workers plausibly add skills when expected returns (e.g., recruiter reach, search visibility) exceed the time costs of updating their profile. A key advantage of my design is that it leverages within-person changes in reported skills between 2019 and 2021, conditioning on rich pre-treatment information. This substantially mitigates cross-sectional concerns: time-invariant tendencies to under-report or over-report basic or widely shared skills and idiosyncratic choices about how granular to list

skills are absorbed by baseline controls for the stock and composition of skills. Identification of effects relies on the change in reported skills rather than level differences in reporting styles. In addition, the presence of current and former colleagues in one's LinkedIn network may exert a form of social pressure that discourages workers from substantially overstating their skills.

A practical measurement issue is strategic reporting some workers may delay updating their profile even after acquiring new skills. In that case, absence of an update does not imply absence of human capital accumulation, and some workers who actually acquired skills will be classified as controls. This misclassification pushes estimates toward zero if learning is benefits career outcomes. The focus on longitudinal updates reduces (but cannot eliminate) this concern by tying measurement to actual profile changes over time.

Profile staleness poses another data quality concern. If workers rarely update their LinkedIn profiles, skills acquired may not be added to profiles. The sample construction limits staleness by requiring complete baseline information in January 2019 and profile presence back to 2015 and through 2023. I also report robustness dropping workers with no recorded position changes from 2015–2023—retaining profiles demonstrably updated at least once. Appendix 5.B.2 shows that results are very similar to main results.

In sum, exploiting longitudinal variation and conditioning on rich pre-treatment histories addresses many cross-sectional reporting issues and reduces misclassification from non-updating. Moreover, there are strong reasons to interpret skill additions as evidence of human-capital accumulation. Additions are discretionary and time-costly. Platform features (connections, endorsements, assessments, and employer visibility) create reputational costs for misreporting. Accordingly, when a worker adds a skill, the update most plausibly reflects newly acquired productive capability, and thus provides a credible proxy for on-the-job human capital accumulation. A separate subsection on threats to identification (Section 5.3.3) discusses whether the timing and nature of skill reporting on LinkedIn could bias return estimates of skill accumulation on career outcomes.

5.2.5 Job Mobility

To measure labor market mobility and its relation with earnings, I focus on positions and 4-digit occupations. Occupation changes imply position changes, while the

reverse is not true. I use these changes to create two measures of mobility.

First, I count the number of positions a worker held between the start of the treatment period in 2019 and the end of my sample period in May 2023. I subtract one from this number to obtain the number of position switches for each worker. Second, to investigate the earnings component of this mobility, I compute the average year-on-year change in predicted earnings for each year throughout the sample period. If a change in position between the previous and the current year goes along with an above-average earnings increase, it is labelled as a switch towards a higher-paying position. On average, workers switch position 0.72 times, 58% of which are towards a higher-paying position.

I apply the same logic to 4-digit occupations and, to this end, rely on occupation level average earnings. I find that workers switch occupations 0.39 times on average, of which 45% are towards higher-paying occupations.

5.2.6 Earnings

I use two sources of earnings data that capture different subsets of the total variation in earnings that is typically contained in administrative data. First, I use predicted earnings from job characteristics such as job title, location, year, and firm provided by Revelio Labs using a proprietary algorithm. Second, I use occupation-level average earnings from the ACS to establish robustness for the results using predicted earnings.

I use gross yearly earnings rather than logs as is common in the literature studying labor market returns to training programs (McCall et al. 2016).

Predicted Earnings

The model to predict earnings is trained on H1B visa applications, job vacancies, and self-reported salaries on websites such as levels.fyi and Glassdoor. Inputs to the prediction are mainly job characteristics such as job title, firm, and location, as well as calendar year and job tenure. Crucially, workers' skills and other characteristics such as gender or degree do not enter the prediction.

By construction, this measure does not capture variation in earnings that is within all these inputs. Therefore, I cannot capture variation in earnings stemming from, e.g., bargaining (Caldwell et al. 2025a) that does not affect any of the inputs to our prediction. However, given that position data is very detailed, I am still able

to capture significant parts of total variation in earnings, especially because job titles give detailed information about career progression. For instance, transitioning from “(junior) accountant” to “staff accountant” to “senior accountant” implies significant increases in predicted earnings within worker across time, even within firm and within narrow occupation category in our data.⁵

To better assess which inputs to the prediction are most related to earnings trajectories, I analyze to what extent year-on-year changes in earnings are related to changes in job title, firm, and metropolitan statistical area (MSA). This clarifies what position characteristics might mediate a potential effect of accumulating skills on predicted earnings.

Table 5.2 shows coefficient estimates and group shares of a regression of workers’ year-on-year change in log predicted earnings on mutually exclusive and exhaustive group indicators for all the possible changes in the position a worker currently holds (see Appendix 5.A.2 for details). From one year to another, workers might change different combinations of job title, firm, and MSA. Column (1) indicates that, on average, changes in job position characteristics are associated with increased earnings, i.e., workers make moves to higher-paying positions. The “No change” row illustrates that with no change in either of the inputs, workers gross yearly earnings increase by roughly 1.5% on average (column (1)). Column (2) indicates that, across my sample period, almost 80% of workers belong to this group. Furthermore, column (3) indicates that, as expected, the contribution to the share of variance in earnings changes the model explains is small at 3.3%.

⁵This is consistent with Deming (2024), who argues that most earnings increases for workers beyond the start of their careers occur within-job instead of between job, with a job being defined as a employer-by-occupation pairing, which, by construction, is much coarser than the definition of a position in our data.

Table 5.2: Decomposition of predicted earnings change by switch of position characteristic

Change from year $t - 1$ to t	β_g (in %)	Share (p_g in %)	Share of model variance (in %)
	(1)	(2)	(3)
Title only	12.365	5.703	72.315
Title + Firm	3.572	8.741	8.954
Title + Metro	10.485	0.932	8.928
Title + Firm + Metro	4.340	4.086	6.493
No change	1.521	79.603	3.304
Firm + Metro	0.423	0.255	0.004
Metro only	0.592	0.027	0.001
Firm only	0.058	0.653	0.000
R^2	0.034		
Observations	40,236,035		

Notes: Column (1) presents predicted group-level annual earnings differences (β_g). Column (2) indicates the share of individuals in the estimation sample belonging to a specific switching group. Column (3) shows the contribution of each group g to the variance of the predicted annual earnings changes, calculated as $p_g \times (1 - p_g) \times (\beta_g)^2$ divided by its sum over all g . The estimated model includes eight period-fixed effects. 2015 - 2016, 2016- 2017, ..., 2022 - 2023.

The second largest group changes job title (as well as potentially firm and metro area), which is true for almost 20% of workers. The average earnings increases from these changes are substantial, and the 6% of workers solely changing job title (first row) see their earnings increase by more than 12% on average. Taking these numbers together, changing just job title carries by far the biggest explanatory power for earnings changes in our data.⁶ Since substantial numbers of workers change job title as well as either firm or MSA, and these changes are associated with large earnings increases on average, they also contribute significantly to the share of variance explained.

Finally, comparatively fewer workers switch location and firm, which also go along with smaller increases in earnings, and therefore contribute less to explaining variance in earnings changes.

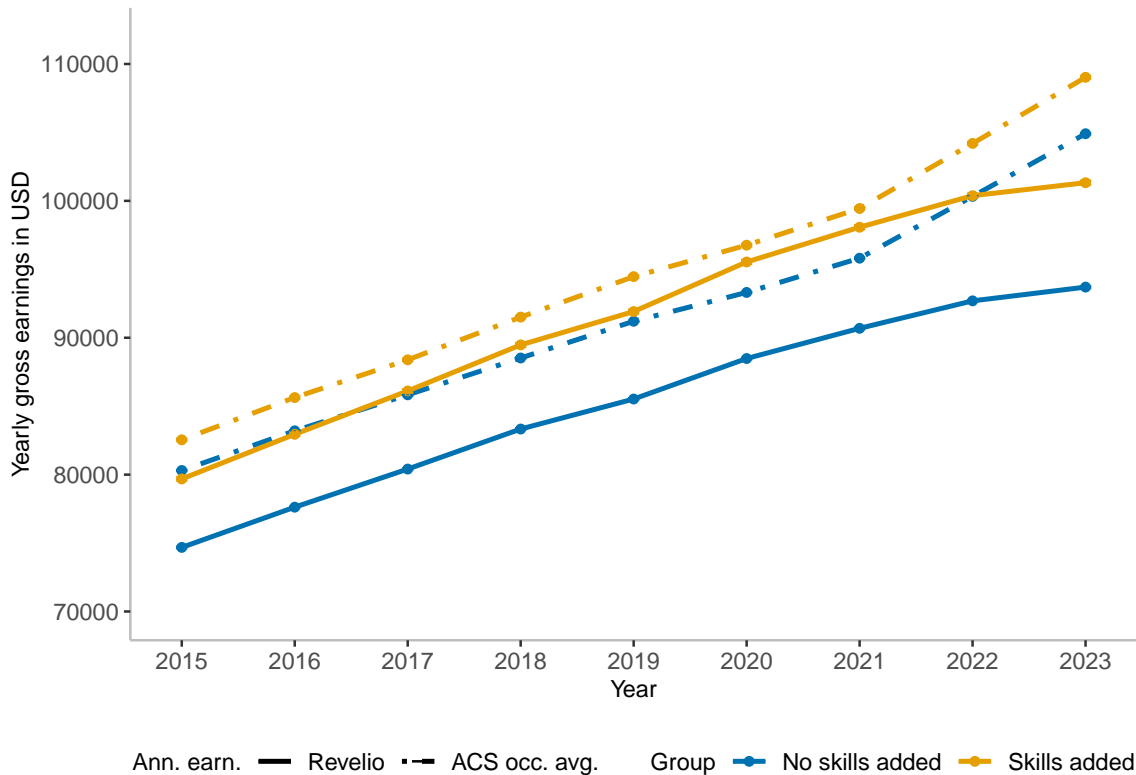
ACS Occupation-Level Average Earnings

To provide robustness, I use occupation level average earnings from the ACS based on 95 4-digit SOC 2018 occupation groups. To this end, I adopt the same criteria for sample inclusion as with the LinkedIn data. By construction, this measure cannot account for within-occupation variation in earnings. However, between-occupation

⁶The R^2 from that regression is rather low at 3.4%. This is because I do not account for the manifold combinations of origin and destination fixed effects for job titles/firms/MSAs in a two-way fixed effects manner as pioneered by Abowd et al. (1999). Rather, I only use switching indicators to explain earnings variation within worker over time since the interest is in identifying what a change in earnings induced by skill accumulation operates through as a mediator.

earnings differentials have been found to explain large shares of wage inequality across workers (e.g., Blau and Kahn 2017).

Figure 5.2: Average predicted and ACS earnings for treatment and control group over time



Notes: Average annual earnings by calendar year and group for predicted earnings (“Revelio”) and occupation level average earnings from the ACS (“ACS occ. avg”).

Figure 5.2 plots average within-worker earnings trajectories over time from both data sources, further distinguishing treatment and control groups. Reassuringly, earnings trends look similar in both data sources, albeit the post-Covid earnings increases are larger in the ACS than Revelio Labs earnings data. In addition, the level of ACS earnings is higher. The latter fact does not bias my estimates since I control for past earnings. The correlations between the occupation average earnings measures from both data range from 0.80 (lowest) in 2020 to 0.83 (highest) in 2015.

5.3 Empirical Framework

5.3.1 Identification Strategy

I use an identification strategy based on a selection-on-observables or unconfoundedness assumption (Rosenbaum and Rubin 1983; Imbens 2004). Simply put, I compare outcomes of treated and control units similar in terms of the full set of pre-treatment covariates and predicted earnings (Imbens and Xu 2024).⁷ Given data on the full pre-treatment education and career trajectory as well as earnings of workers, I identify the causal effect of on-the-job skill accumulation on earnings by comparing workers with almost identical pre-treatment resumes and earnings trajectories.

Treatment is defined as follows

$$D_i = \begin{cases} 1 & \text{if } \Delta_{19-21} \# \text{ skills}_i > 0, \\ 0 & \text{if } \Delta_{19-21} \# \text{ skills}_i = 0. \end{cases}$$

That is, workers are in the treatment group if they added one or more skill(s) to their profiles between 2019 and 2021.

Of course, this is a very coarse coding in that only the extensive margin of skill accumulation is captured rather than the number of added skills. While most workers add skills only once and, in addition, more than 60% merely add a single skill, a substantial remainder adds more skills (Table 5.1). Furthermore, I do not distinguish workers by the type of skill(s) added. Collapsing this heterogeneity into a binary indicator induces measurement error in skill accumulation. However, given that only 17.6% of workers in my sample add skills, there is a trade-off between accounting for heterogeneity in treatment and statistical power. Section 5.5.3 probes the robustness of my findings to capturing the intensive margin of skill accumulation in addition.

For point estimates to have a causal interpretation, two assumptions need to be satisfied. First, a conditional independence of treated and untreated potential

⁷An alternative identification strategy is to rely on within-person within-year variation in a two-way-fixed-effects approach. The fundamental difference between the two is that the two-way-fixed-effects approach does not adjust for differences between treated and control group in lagged outcomes and other pre-treatment covariates (Arkhangelsky and Imbens 2024). Instead, it assumes that treated and control units differ in unobservables that are correlated with a time-invariant determinant of the outcome, the person fixed effect. Unconfoundedness, by contrast, is satisfied if selection is purely based on past rather than future outcomes. For a setting like mine, it is probably less plausible to assume that potential confounders are time-invariant than to believe adding skills is unrelated to potential outcomes conditional on a vector of past earnings and other pre-treatment characteristics (Angrist and Pischke 2009).

outcomes from the treatment conditional on pre-treatment covariates and outcomes. Second, the average treatment effect needs to be estimated at every value for pre-treatment covariates and outcomes. In other words, I require treatment and control samples to have overlapping distributions of propensity scores.

Let Y_{it} denote some career outcome for individual i in May of year t with $t = 2015, 2016, \dots, 2023$. In terms of notation, define $T+ \equiv \{2021, 2022, 2023\}$ as including all post-treatment periods. $\mathbf{Y}_{i,LAG}$ collects pre-treatment outcomes. In addition, let pre-treatment covariates be denoted by \mathbf{X}_i .

Assumption 1 (Unconfoundedness)

$$D_i \perp\!\!\!\perp (Y_{iT+}(0), Y_{iT+}(1)) \mid \mathbf{Y}_{i,LAG}, \mathbf{X}_i$$

Assumption 2 (Overlap)

$$0 < \Pr(D_i = 1 \mid \mathbf{Y}_{i,LAG}, \mathbf{X}_i) < 1.$$

In words, assumption 1 posits that conditional on pre-treatment covariates and outcomes, whether or not a worker gets to acquire some skill is unrelated to potential outcomes.

The set of covariates \mathbf{X}_i tries to capture workers full career trajectories before the start of the treatment period, akin to what is usually contained in a resume. In particular, the set includes the month in which the profile was first scraped, age and gender in 2019, fixed effects for five highest degrees obtained, 22 fields of study, and five college ranking brackets. Furthermore, it includes the ratio of actual to potential work experience, fixed effects for the numbers of years of job tenure, for 22 occupation categories, and for 154 different MSAs (all in 2019). Additionally, fixed effects for 22 occupation categories at the start of workers careers and the graduation cohort for the first tertiary degree are included. Finally, the number of job switches between 2015 and 2019 capture sorting and search behavior in the labor market pre-treatment.⁸ Finally, information about workers pre-treatment skill set is captured by including the number of reported skills as well as what share of skills falls into the mutually exclusive categories of general, managerial, and occupation-specific skills.

Lagged outcomes $\mathbf{Y}_{i,LAG}$ serve both as strong predictors of future earnings as well as proxies for underlying ability and labor market attachment. The information

⁸In regressions using post-treatment mobility between positions and occupations, this variable is a lagged outcome rather than pre-treatment covariate.

contained in lagged outcomes and pre-treatment characteristics captures both predictors of treatment and potential confounders that affect career outcomes. I exclude post-treatment variables to avoid bias from conditioning on potential mediators or colliders.

Assumptions 1 and 2 identify the average treatment effect (ATE), which is the causal estimand of interest in this paper.

5.3.2 Assessing Identification

I assess the validity of my identification strategy using placebo tests and sensitivity analyses (Oster 2019; Cinelli and Hazlett 2020).

While assumption 1 is not directly testable, it can be bolstered using placebo tests. These refer to estimating treatment effects that are known to be zero. More specifically, one of the pre-treatment controls is used as a pseudo-outcome that serves as a proxy for the target outcome while still conditioning on the full set of remaining pre-treatment controls (Imbens and Xu 2024). In this case, lagged earnings are particularly suitable, since they should be a good proxy for current and future earnings.

Another way to mitigate concerns regarding potential confounders is to use sensitivity analyses. These analyses quantify how strongly the relationship between unobserved confounders, the treatment and the outcome would need to be to overturn treatment effect estimates. In addition, these relationships can be expressed relative to the strength of already observed control variables.

5.3.3 Threats to Identification

Reverse Causality from Timing of Skill Additions

The primary threat to identifying a causal effect of on-the-job human capital accumulation on earnings is reverse causality. Due to the rather long period in which workers could potentially add skills to their profiles (Jan 2019 to May 2021), it is not guaranteed that job switches are always preceded by skill additions.⁹ In

⁹In other words, treatment occurs at different times for different workers. Such a setting lends itself to event-study designs in relative time. Such approaches have been adopted frequently in the literature evaluating active labor market programs (e.g., Biewen et al. 2014). However, those require two strong assumptions: (i) accurate measurement of the timing of skill acquisition, and (ii) behavioral consistency in how workers update their profiles. Since skill additions are only timestamped when scraped from profiles (not necessarily when skills were acquired), and update behavior is heterogeneous and unobservable, I prefer a simpler approach that avoids relying on

fact, workers might switch to a higher-paying job, master new skills there and update their profile accordingly. In this case, the increased earnings that preceded human capital accumulation would wrongly be attributed to the latter and show up as part of the estimated treatment effect.

To avoid this, one could easily drop workers who add skills only after having a started a new position. However, this is far from perfect. First, workers might only update their profiles and skills only after having started a new position. Since I have timestamps of when positions were started but not when skills were acquired but rather only when skills were added to profiles, it could be that skills added after a new position had been started were actually accumulated during the previous job and helped secure the new job. The time window to determine what skills belong to what position crucially depends on the assumption one makes about the nature of human capital accumulation on the job in addition to how workers update their profiles. Depending on the type of skill, it might take different amounts of time to master it. For instance, people management skills might take longer to acquire than, say, programming skills for a specific language, which could possibly be mastered during a full-time bootcamp lasting for a few weeks. Considering all of this, I consider skills added to profiles more than 3 months after having started a new position to have been mastered in that new job rather than the previously held one. While the three-month threshold is necessarily imperfect, it strikes a balance: short enough to exclude clear cases of reverse causality, but long enough to allow for realistic delays in profile updates.

However, skills added after having started a new position might facilitate subsequent job switches. Potential earnings increases from those should be part of the estimated treatment effect on later post-treatment outcomes. Therefore, dropping all workers having added skills 3 months or more after starting a new position confounds the treatment effect as well in that it removes substantial parts of the identifying variation.

Taking all of these considerations into account, I drop workers from the treatment group who meet the following two conditions for a robustness check (see section 5.5.3): (i) they switched jobs at least once during the treatment period (2019 to May 2021), and (ii) they added skills three months or more after starting their most recent position as observed in the May 2023 snapshot. This restriction removes a

specific assumptions about timing of skill additions.

subset of treated workers where reverse causality is most concerning. In particular, I drop 87,172 workers or 10.7% of the treatment group.

Strategic Reporting

A remaining threat is omitted variable bias from strategic reporting. Workers may update profiles mainly when job hunting or anticipating a move, adding skills they already possessed. Such time-varying search intensity can raise both the probability of adding skills (i.e., being in the treatment group) as well as subsequent career outcomes including mobility and earnings. The latter effect might operate through applications or recruiter contact. If so, estimated effects may load partly on this transitory shock rather than human-capital growth per se.

Net bias from job hunting is a priori ambiguous because two forces operate in opposite directions. On one hand, search intensity is an omitted variable positively correlated with both adding skills and mobility/earnings, which biases estimates upward. On the other, not all workers who learn new skills immediately report them. Some workers who actually acquired skills remain in the control group. This attenuates effects toward zero assuming that skill accumulation benefits career outcomes.

Rich pre-treatment controls—including lagged earnings, baseline skill portfolios, and detailed mobility histories from 2015–2019—should net out time-invariant differences in job-search propensity and career satisfaction between treated and control workers. Indeed, placebo regressions on pre-2019 outcomes show little evidence of pre-trend (see Section 5.5.1). However, these checks cannot rule out transitory shocks concentrated in 2019–2021. Sensitivity analyses indicate that a sufficiently strong omitted factor jointly tied to adding skills and earnings could attenuate or overturn estimates.

5.3.4 Empirical Strategy

Estimating Propensity Scores

I estimate propensity scores using a probability forest estimator (Athey et al. 2019), which flexibly captures nonlinearities and high-dimensional interactions between covariates without the need for parametric assumptions. I use 750 trees

for estimation.¹⁰ Formally, I estimate the propensity score

$$e(\mathbf{Y}_{i,\text{LAG}}, \mathbf{X}_i) = \mathbb{P}(D_i = 1 \mid \mathbf{Y}_{i,\text{LAG}}, \mathbf{X}_i)$$

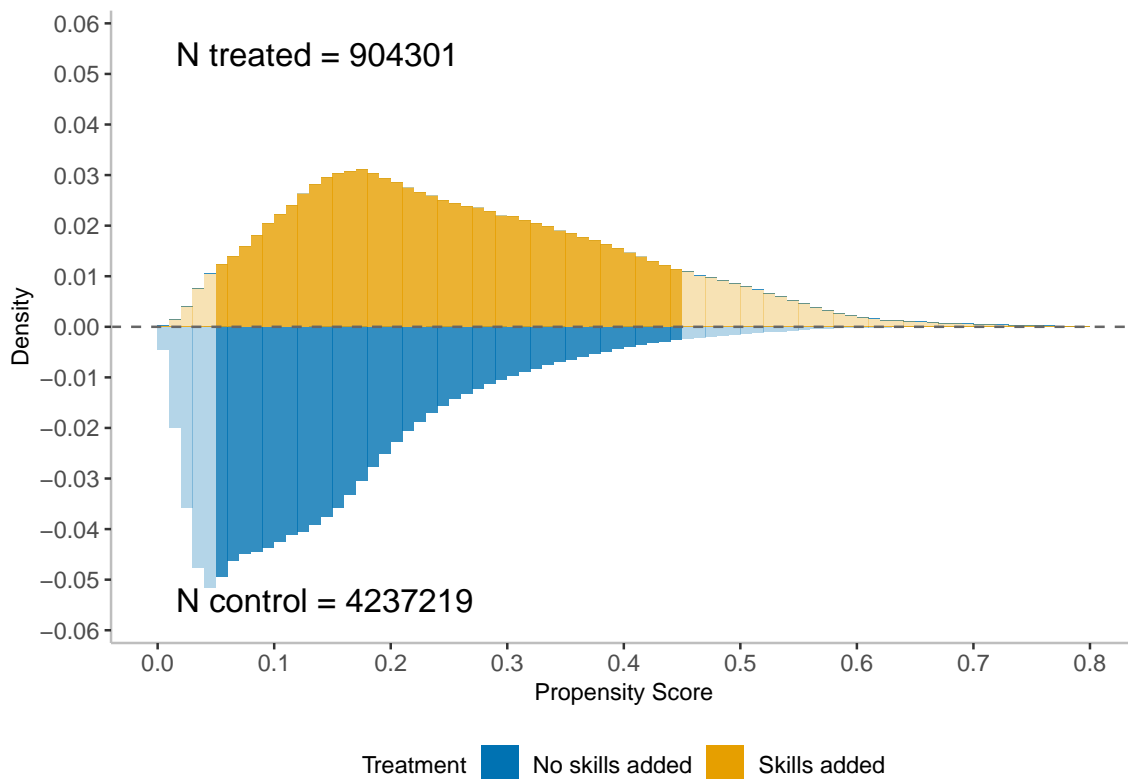
as the conditional probability that individual i adds at least one skill between 2019 and 2021, given lagged outcomes and other pre-treatment covariates. The use of a flexible machine learning estimator is especially attractive in this prediction setting, since treatment assignment may be driven by complex, nonlinear combinations of prior characteristics (e.g., past earnings trajectories, skill levels, or educational background).

Because the probability of treatment is relatively low in the sample—only 17.6 percent of individuals add a skill—the estimated propensity score distribution is skewed, with the vast majority of scores lying below 0.5. To ensure sufficient common support between treated and control units, I drop observations with propensity scores outside the interval $[0.05, 0.45]$, which corresponds to the region where both groups have substantial representation. This approach follows the common practice of restricting analysis to the region of overlapping support (Imbens and Rubin 2015) and avoids relying on treated (or control) units for which no comparable counterparts exist.

While it is less aggressive than the “[0.1, 0.9]” suggested in Crump et al. (2009), the cutoffs reflect the empirical distribution in a rare-treatment setting, where lack of overlap arises primarily from very small propensities. My approach is consistent with recent proposals that down-weight tails of the propensity score to improve finite-sample performance (e.g., matching weights proposed in Li et al. 2018). I demonstrate that using $[0.05, 0.45]$ yields (i) substantial overlap in treated vs. control propensity distributions, (ii) a large effective sample size. My approach therefore balances the desire to maintain a large, representative sample with the need to avoid extrapolation into regions of poor support.

Figure 5.3 verifies the distributional overlap visually and assesses covariate balance conditional on the estimated scores. The opaque parts of the histogram correspond to trimmed observations. In the subsequent analysis, I use the trimmed sample to estimate average treatment effects.

¹⁰While the package defaults to 4000 trees, this is computationally prohibitively expensive in my case. Athey et al. 2019 note that for the estimation of conditional means as in my case, fewer trees are sufficient to ensure good performance.

Figure 5.3: Propensity score distributions of trimmed sample

Notes: Histograms of the propensity scores (\hat{e}) estimated through GRF (Athey et al. 2019) are depicted for treatment and control group. Shaded areas indicate trimmed units with $\hat{e} < .05$ or $\hat{e} > .45$ that are removed from the analysis sample. Observation numbers refer to the trimmed sample.

Table 5.3 displays treatment and control means in pre-treatment covariates for the trimmed sample. This exercise is informative about the variables that should be accounted for to make the unconfoundedness assumption more plausible and therefore to bolster the causal interpretation of the subsequently estimated effects.

As expected, there are significant mean differences (both statistically and economically) between the groups in most of the pre-treatment covariates shown. More precisely, the group of workers adding skills is much less female, slightly older and in terms of educational degrees and prestige of institutions similar compared to the control group not adding skills. Furthermore, the treatment group has slightly higher labor market attachment and higher job mobility. Finally, treated workers reported higher shares of managerial and specific skills as well as a substantially higher number of skills pre-treatment.

Figure 5.2 shows that pre-treatment predicted and ACS occupation level average earnings are also highly correlated with treatment assignment.

Table 5.3: Balancing table after trimming

	No skills added		Skills added		Mean diff.
	Mean	Std. Dev.	Mean	Std. Dev.	
Female	0.44	0.50	0.38	0.49	-0.06***
Age	38.32	9.96	38.54	9.98	0.22***
Associate	0.07	0.26	0.07	0.26	-0.00*
Bachelor	0.56	0.50	0.57	0.49	0.02***
Master	0.29	0.45	0.29	0.45	0.00***
Professional Degree	0.04	0.20	0.03	0.17	-0.01***
Doctorate	0.04	0.21	0.04	0.19	-0.01***
College ranking	4.62	1.37	4.59	1.37	-0.03***
Actual/potential experience	0.89	0.14	0.90	0.13	0.01***
# of positions (2015–2019)	1.97	1.01	2.16	1.05	0.19***
# of skills	19.04	10.13	23.24	11.05	4.21***
Skill share (general)	0.53	0.32	0.47	0.31	-0.06***
Skill share (managerial)	0.23	0.25	0.24	0.25	0.01***
Skill share (specific)	0.24	0.31	0.29	0.31	0.04***
Earnings 2015	74,682	40,820	79,684	43,305	5,003***
Earnings 2016	77,622	41,877	82,945	44,154	5,322***
Earnings 2017	80,406	43,012	86,113	45,112	5,707***
Earnings 2018	83,331	44,394	89,478	46,383	6,147***
Earnings 2019	85,523	45,526	91,914	47,275	6,392***
Observations	4,237,219		904,301		5,141,520

Notes: Means and standard deviations of characteristics for treatment and control groups. Stars indicate p -values from a t -test of differences in means: *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$.

Estimating Outcomes

Identifying the average career return to on-the-job skill accumulation relies on the comparison of workers with almost identical pre-treatment resumes and earnings trajectories that differ only with respect to whether they acquired skills on the job or not. This gives rise to a selection-on-observables approach that aims at netting out confounders that affect both treatment assignment and post-treatment job mobility and earnings. I am therefore interested in the coefficient τ of the following regression

$$Y_{iT+} = \beta_0 + \mathbf{Y}'_{i,LAG} \boldsymbol{\delta} + \tau \cdot D_i + \mathbf{X}'_i \boldsymbol{\beta} + \varepsilon_i, \quad (5.1)$$

where $\mathbf{Y}_{i,LAG}$ is a vector of lagged outcomes, D_i is a dummy indicating that a worker added skills between 2019 and May 2021, and \mathbf{X}_i contains pre-treatment covariates, some of which are included linearly (age in years, gender, ratio of actual to potential

work experience, number of skills, skill shares for general, managerial, specific (all in 2019) and others as fixed effects (highest degree, field of study, college ranking brackets, years of job tenure, occupations, MSA, number of positions since 2015 (all in 2019), 22 occupations at career start, graduation cohort). ε_i is an error term.

Standard errors are robust for predicted earnings outcomes and clustered at the occupation level for occupation level average earnings from the ACS.¹¹

Imbens and Xu (2024) emphasize that ensuring adequate overlap—often via trimming—is more consequential than the specific estimator used. In empirical re-analyses with good overlap, regression adjustment, inverse-probability weighting, doubly robust estimators, and modern machine-learning implementations yield very similar treatment effect estimates. This motivates my focus on overlap diagnostics and trimming. I report OLS with rich controls as a transparent baseline.

5.4 Results

5.4.1 Job Mobility

Table 5.4: OLS estimates for the effect of accumulating skills on job mobility

Dependent variable (betw. 2019–2023)	# of position switches (1)	# of switches to higher-paying pos. (2)	# of occupation switches (3)	# of switches to higher-paying occ. (4)
Skills added	0.078 (0.005)	0.049 (0.003)	0.025 (0.006)	0.016 (0.004)
Control mean	0.696	0.404	0.381	0.188
Treated - control mean	0.182	0.107	0.069	0.037
R^2	0.199	0.164	0.126	0.106

Notes: $N = 5,141,520$. OLS regressions of career outcomes on treatment indicator with the following controls and fixed effects. Fixed effects: five highest degrees obtained, 39 fields of study, five college ranking brackets, years of job tenure, 22 occupations, 154 MSAs, number of positions (all in 2019), month of initial scrape, 22 occupations at career start, graduation cohort. Controls: age in years, gender, ratio of actual to potential work experience, number of skills, skill shares (general, managerial, specific), lagged outcomes from 2015 to 2019. Standard errors in parentheses are robust in columns (1), (2) and clustered at the 2021 occupation level for outcomes (3), (4).

Table 5.4 shows estimates of the effect of on-the-job skill accumulation on job mobility outcomes upon including pre-treatment covariates as well as lagged earnings and job mobility. Column (1) shows that workers accumulating skills are more mobile with respect to job positions. In fact, accumulating skills increases

¹¹The number of occupations or rather clusters is 95 and therefore exceeds often-cited thresholds for asymptotic validity of clustered standard errors, rendering small-sample adjustments unnecessary.

the number of times a worker switches position between 2019 and 2023 by 0.078 or 11.2% relative to the control mean. However, even though job mobility has been shown to be an important part of workers' earnings growth (Topel and Ward 1992; Deming 2024), column (1) merely speaks to mobility itself, some of which need not be related to earnings increases (Hall and Mueller 2018; Adda and Dustmann 2023). Combining mobility with the direction of earnings changes, column (2) shows that accumulating skills enhances workers' mobility that is geared towards higher-paying positions. More specifically, accumulating skills increases the number of switches to higher-paying positions by 0.049 or 11.9% of the control mean.

Turning to occupations, the pattern demonstrated with job positions is confirmed: skill accumulation on the job increases the number of occupations switches as well. In particular, treated workers switch occupations 0.025 times or 6.6% more often than control group workers. While earnings effects of occupational mobility are ex-ante unclear (Groes et al. 2015), accumulating skills might help transition to better-paying occupations. Column (4) demonstrates that this hypothesis is borne out in the data. In particular, workers accumulating skills switch to higher-paying occupations 0.016 times more than workers not adding skills. With regards to the control mean, this effect corresponds to an increase of 8.5%. Given that occupation switches are a subset of position switches, effect sizes for occupations must be smaller than for positions in absolute magnitude.

5.4.2 Earnings

This section provides estimates of the effect of skill accumulation on the job on subsequent earnings. This helps quantify the earnings effects that go along with increased mobility reported in section 5.4.1.

Table 5.5 shows results of estimating equation (5.1) for predicted earnings in 2021, 2022, and 2023. More precisely, accumulating skills between 2019 and 2021 yields an immediate return to May 2021 predicted earnings of 650\$ or 0.72% of the mean earnings among workers not accumulating skills. In addition, returns to predicted earnings in 2022 (2023) amount to 871\$ (882\$) or 0.94% (0.94%) with respect to the control mean. Earnings returns in 2022 and 2023 are significantly larger than in 2021, suggesting that earnings returns increase over time rather than fading out. R^2 values decrease as outcomes move farther from the end of the treatment window, presumably because idiosyncratic shocks to earnings accumulate.

Table 5.5: OLS estimates for the effect of accumulating skills on predicted earnings

Dependent variable	Earnings 2021	Earnings 2022	Earnings 2023
Skills added	650 (33)	871 (49)	882 (46)
Control mean	90,689	92,695	93,699
Treated - control mean	7,321	7,619	7,561
R^2	0.744	0.678	0.644

Notes: $N = 5,141,520$. OLS regressions of predicted earnings on a treatment indicator with controls and fixed effects in addition. Fixed effects: initial scrape month, 5 highest degrees obtained, 39 fields of study, 5 college ranking brackets, years of job tenure, 22 occupations, 154 MSAs, number of positions since 2015 (all in 2019), 22 occupations at career start, graduation cohort. Controls: age in years, gender, ratio of actual to potential work experience, number of skills, skill shares for general, managerial, specific (all in 2019), lagged earnings from 2015 to 2019. Robust standard errors in parentheses.

Table 5.6 repeats this exercise leveraging occupation level average earnings from the ACS. Since these capture only between occupation variation in earnings, they vary within-worker only upon a change in occupation. Unsurprisingly, return estimates are much smaller in absolute terms and provide a lower bound to the actual effect. Return estimates to on-the-job skill accumulation amount to 328\$ or 0.34% of the control mean. In 2022 (2023), returns to predicted earnings increase to 0.54% (0.61%) of the control means or 545\$ (635\$) in absolute terms.

Reassuringly, estimates are smaller but in the same direction, lending further credibility to the estimates relying on predicted earnings. In addition, the pattern of increasing rather than fading returns is borne out between occupation as well.

Table 5.6: OLS estimates for the effect of accumulating skills on ACS earnings

Dependent variable	ACS Earnings 2021	ACS Earnings 2022	ACS Earnings 2023
Skills added	328 (151)	545 (256)	635 (299)
Control mean	95,813	100,326	104,905
Treated - control mean	3,603	3,842	4,092
R^2	0.731	0.674	0.645

Notes: $N = 5,141,520$. OLS regressions of predicted earnings on a treatment indicator with controls and fixed effects. Fixed effects: initial scrape month, 5 highest degrees obtained, 39 fields of study, 5 college ranking brackets, years of job tenure, 22 occupations, 154 MSAs, number of positions since 2015 (all in 2019), 22 occupations at career start, graduation cohort. Controls: age in years, gender, ratio of actual to potential work experience, number of skills, skill shares for general, managerial, specific (all in 2019), lagged earnings from 2015 to 2019. Standard errors in parentheses are clustered at the occupation level of the respective year.

5.5 Robustness

5.5.1 Placebo

Table 5.7 displays results from estimating equation 5.1 with lagged 2019 earnings as the dependent variable. Other than moving lagged 2019 earnings from the left-hand side to the right-hand side of the equation, it is the same specification as used in Tables 5.5 and 5.6. This causal effect is known to be zero: Skill accumulation can only happen after the date at which these earnings were recorded, rendering it impossible for them to affect these lagged earnings.

Table 5.7: OLS placebo estimates for the effect of accumulating skills on lagged earnings

Dependent variable	Earnings 2019 (1)	Acs Earnings 2019 (2)
Skills added	74 (34)	62 (137)
Control mean	85,523	91,204
Treated - control mean	6,358	3,261
R^2	0.883	0.839

Notes: $N = 5,141,520$. OLS regressions of predicted/occupation level average earnings on a treatment indicator with the following controls and fixed effects. Fixed effects: 5 highest degrees obtained, 39 fields of study, 5 college ranking brackets, years of job tenure, 22 occupations, 154 MSAs, number of positions since 2015 (all in 2019), 22 occupations at career start, graduation cohort. Controls: age in years, gender, ratio of actual to potential work experience, number of skills, skill shares for general, managerial, specific (all in 2019), lagged earnings from 2015 to 2018. Robust (1)/clustered at the occupation level (2) standard errors in parentheses.

Indeed, column (1) shows that adding one or more skill(s) between 2019 and 2021 only yields a very modest earnings gain of 74\$ or 0.087% of the control mean. While not exactly zero, the estimate is roughly an order of magnitude smaller than the estimates on post-treatment earnings and only borderline significant at the 5% level. There is a similar pattern for ACS occupation level average earnings in column (2), where the effect using all controls is similar in absolute size, but is statistically not significantly different from zero.

While this near-null effect estimate of accumulating skills on pre-treatment earnings does not rule out violations of the unconfoundedness assumption, it confirms a conditional independence that is implied by it (Imbens and Xu 2024). The small placebo effect estimate therefore provides strong support for

the unconfoundedness assumption and lends further credibility to the estimated effects on post-treatment outcomes.

5.5.2 Sensitivity

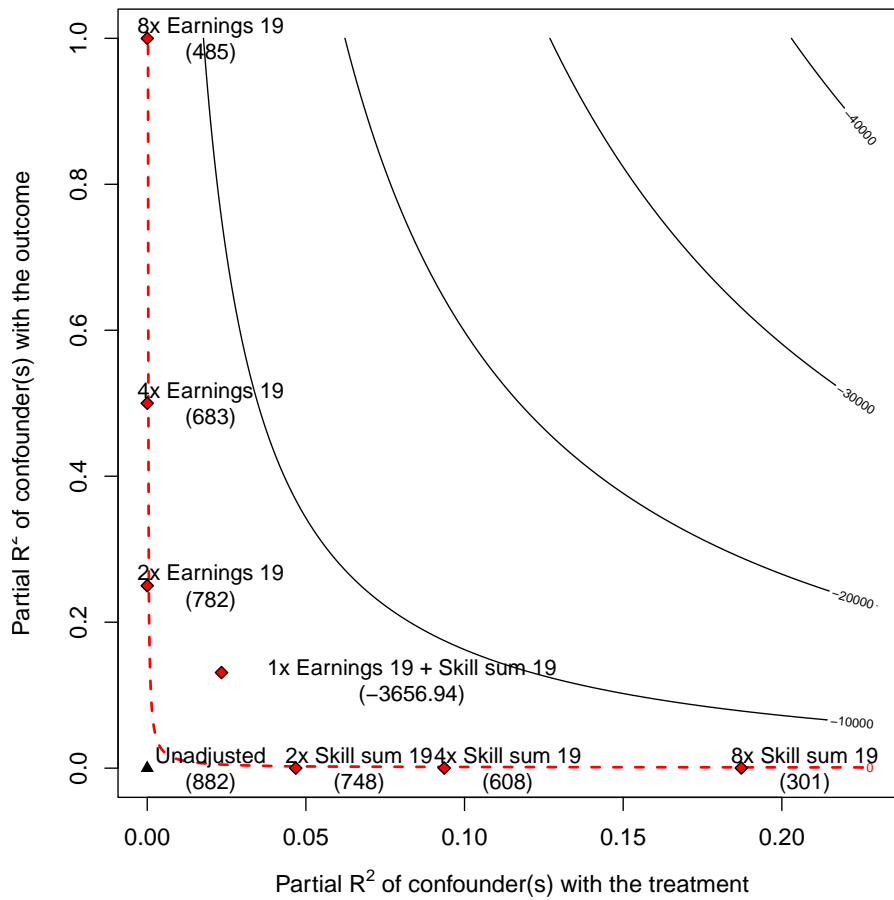
Sensitivity analyses assume that unconfoundedness (Assumption 1) only holds if observed pre-treatment covariates and lagged outcomes (short: pre-treatment information) are supplemented with a previously unobserved confounder Z . In other words, they probe how much an omitted factor would have to matter—in terms of its remaining association with both treatment (“Skills added”) and 2023 earnings after conditioning on all pre-treatment information—to alter the results. The idea is simple: if the estimate stays positive unless the unobserved confounder is very strongly tied to both treatment assignment and the residual part of earnings, then the result is robust (see, e.g., Oster 2019; Cinelli and Hazlett 2020; Imbens and Xu 2024).

To make this concrete, I add benchmarks that scale the strength of Z relative to two components of pre-treatment information in equation 5.1: lagged earnings in 2019 and the number of reported skills in 2019 (e.g., “2× Earnings 2019” means Z is twice as predictive of the residual outcome as 2019 earnings are, conditional on the controls). These two variables display the strongest partial correlations with the outcome and with treatment, respectively.¹² They are therefore reasonable upper-bound benchmarks for what an omitted factor might achieve.

Using 2023 earnings as the outcome, Figure 5.4 maps out, for a wide range of hypothetical omitted-factor strengths, what the treatment coefficient would become after “accounting” for such a factor. The horizontal axis measures how tightly Z would still be related to treatment net of pre-treatment information; the vertical axis measures how tightly Z would still be related to 2023 earnings, again net of pre-treatment information. Contours give the implied treatment effect.

¹²To be precise, I ran two regressions using the same set of controls as in the “Propensity score” specification of Table 5.5, one with 2023 earnings and the other with the treatment indicator as the outcome.

Figure 5.4: Sensitivity analysis for effect on 2023 predicted earnings



Notes: Contour plot for the treatment effect coefficient based on Cinelli and Hazlett (2020).

Along both axes, even large omitted-factor dependence—several times stronger than either lagged earnings or the baseline number of reported skills on its own—attenuates but does not overturn the positive effect on 2023 earnings. To drive the estimate to zero or negative, Z would need to be simultaneously very strongly related to both treatment and residual earnings—roughly akin to combining the predictive content of both benchmarks at once.

Given that the model already explains a substantial share of variation in 2023 earnings (about $R^2 \approx 0.64$; see Table 5.5), it is hard to conceive of a single omitted factor that is this predictive and this aligned with treatment after conditioning on the rich pre-treatment information included. While such a confounder is unlikely, I cannot rule out the presence of time-varying shocks (e.g., career aspirations or employer-provided training) that jointly affect skill acquisition and earnings. Overall, I view the results as moderately robust: they are stable to strong one-sided

confounding (i.e., large dependence with either treatment or outcome alone) but would require an unusually powerful joint confounder to be overturned.

5.5.3 Further Robustness Checks

This section explores the robustness of the previously reported results upon employing a different definition of the treatment as well as implementing additional sample restrictions. In particular, the main results only distinguish workers along the extensive margin of skill accumulation. Treatment is defined based on whether workers added skills or not, irrespective of the number of added skills. Therefore, the first robustness check investigates the intensive margin of skill accumulation using three additional estimating equations. Results are reported in the Appendix. The second set of robustness checks involves dropping profiles that are (i) potentially stale, (ii) have added skills after the onset of the Covid-19 pandemic in February 2020, and (iii) potentially suffer from reverse causality in the sense that they have only added skills after they had started their most recent position.

Intensive Margin of Skill Additions

To probe the robustness of the results with regards to the definition of treatment assignment, I decompose the treatment effect into extensive and intensive margin components. Motivated by the distribution of added skills among treated workers with a mean of 3, a median of 1, and with 62.3% adding exactly 1 skill, 14.0% adding 2, 11.2% adding 3–5, and 12.5% adding 6+ (= 904,301 out of $N = 5,141,520$ workers add skills), I re-define treatment to incorporate the intensive margin in three ways.

First, I replace the dummy for adding at least one skill(s) (D_i) with the raw count of added skills. This specification targets a different estimand: the average return to one additional skill across a zero-inflated, highly skewed distribution of the number of added skills. When most treated workers add exactly one skill and returns are threshold-like, a single linear “per-skill” slope will be mechanically small even if the extensive margin effect is sizable.

Table 5.B.1 shows that if the same controls and fixed effects as in the main earnings results are used, the coefficient on the number of skills added is small and positive (about \$60–\$70 per additional skill across years), while the extensive-margin specification shows earnings gains between 650\$ and almost almost 900\$ for adding any skill (Table 5.5).

To investigate this further, the second specification (equation 5.2) includes both intensive and extensive margin components of the effect of adding skills explicitly:

$$Y_{iT+} = \beta_0 + \mathbf{Y}'_{i,LAG} \boldsymbol{\delta} + \tau \cdot D_i + \gamma D_i \times (\Delta\#of\ skills_i - 1) + \mathbf{X}'_i \boldsymbol{\beta} + \varepsilon_i. \quad (5.2)$$

The total effect of adding skills for worker i is $\tau + \gamma \times (\Delta\#of\ skills_i - 1)$. Table 5.B.4 indicates that estimates of adding exactly one skill (τ) in equation 5.2 are similar to those for adding any number of skills (see Table 5.5) and that the effect of adding additional skills (γ) is smaller than but close to zero. This implies that earnings gains are concentrated on the extensive margin: the total effect is roughly equal to the effect of adding exactly one skill. Therefore, adding additional skills beyond one single skill is not associated with further earnings increases. If anything, this effect is negative.

Finally, I run a specification that replaces the treatment indicator with mutually exclusive dummies for exactly 1, 2, 3–5, and 6+ added skills (reference = 0). Consistent with the previous results, Table 5.B.7 shows that relative to not adding skills, the largest gains accrue at *one to two* added skills. Beyond that, effects are flat or decline: adding *three to five* skills yields similar or slightly smaller gains, and *six or more* does not raise earnings further. The profile is hump-shaped at low numbers of added skills rather than linear. The pattern is the same for ACS occupation level average earnings (see Tables 5.B.2 and 5.B.5) as well as for mobility outcomes (see Tables 5.B.3 and 5.B.6).

Taken together, these analyses indicate that returns to skill accumulation are concentrated on the extensive margin. This pattern is consistent with a nonlinear, step-function relationship between the number of added skills and earnings. A plausible explanation is measurement and reporting noise on the intensive margin: because many workers update their profile only once, that update may reflect a single underlying learning event, while the exact number of skills entered at that time is idiosyncratic (how finely a worker lists the same capability). In that case, counting skills absorbs substantial noise, attenuating the per-skill slope, whereas the adoption indicator still captures the economically relevant decision to update. Overall, the evidence points to an extensive-margin channel—earnings respond to initiating an update, with little incremental payoff from piling on additional skills beyond the first.

Additional Sample Restrictions

Staleness. As detailed in section 5.2.4, I drop profiles with potentially stale positions information from the sample. In fact, 28.8% of the sample have not reported any change to their positions data between 2015 and 2023. While this might accurately reflect workers' career trajectories, it might also partly be driven by staleness. Even though fewer profiles of treated workers exhibit staleness (21% vs. 30%), results are largely unchanged. While treatment effect estimates for both sources of earnings data are slightly larger (Tables 5.B.8 and 5.B.9), those for job mobility outcomes turn out slightly smaller than in the main results (Table 5.B.10). The latter pattern is partly mechanical because the new sample conditions on job mobility at least once during the sample period. Since that condition dropped more workers in the control than in the treatment group, effects are attenuated. Reassuringly and in line with the main results, placebo estimates in Table 5.B.11 are close to zero.

Covid-19. Since the Covid-19 pandemic has caused major upheaval on the labor market, I exclude all workers who have added skills later than January 31, 2020 from this robustness analysis. Unfortunately, this affects 70% of all workers adding skills. Reassuringly, the direction of the estimated effects persists. While estimated effects on earning outcomes are substantially smaller (Tables 5.B.12 and 5.B.13), estimated effects on job mobility are even larger (Table 5.B.14). Standard errors are substantially larger, but conclusions from the main results persist.

Reverse Causality. As detailed in section 5.3.3, I remove treated workers who simultaneously (i) switched jobs at least once during the treatment period, and (ii) added skills three months or more after starting their most recent position as observed in the May 2023 snapshot. Estimated treatment effects on earnings (Tables 5.B.15 and 5.B.16) turn out slightly smaller than the main results. Regarding job mobility (Table 5.B.17), effects on switches to (higher-paying) positions are slightly larger, whereas the corresponding effects on occupations are slightly smaller when compared to the main results. Crucially, placebo effects are still close to zero, albeit slightly larger than in the main results (Table 5.B.18).

5.6 Summary and Concluding Remarks

Longitudinal variation in 5.1 million workers' skill sets on their LinkedIn profiles between 2019 and May 2021 is used to estimate career returns to accumulating

skills on the job. Results demonstrate that accumulating skills benefits labor market mobility and earnings. In particular, workers accumulating skills switch job position 11% and occupation 7% more often than comparable workers who do not accumulate skills. These switches are disproportionately directed towards positions and occupations with higher pay. The annual increase in predicted earnings is almost \$900 in 2023, or close to 1% of earnings of comparable workers who did not accumulate skills. Using occupation level average earnings, increases amount to roughly 0.6% in 2023.

The estimates are derived from comparing career trajectories of workers who differ in whether they accumulated skills between 2019 and 2021 but had otherwise identical prior resumes and earnings. Conditional on this extensive information about career trajectories, the presence of unobserved factors jointly determining skill acquisition and earnings is rather unlikely. Small and barely significant placebo effects on lagged earnings bolster the credibility of the presented estimates. Furthermore, sensitivity analyses suggest that effect estimates are moderately robust to unobserved confounding.

While providing novel evidence on the career returns to skill accumulation using unusually rich data, my approach comes with several limitations. First, even conditional on detailed information about career trajectories, reporting skills is not random and might still be related to the untreated career outcomes of interest. In other words, unobserved confounding cannot be ruled out. Second, although skill additions on LinkedIn profiles are plausibly related to human capital accumulation, they are a proxy for the underlying concept of interest. In addition, the data are not informative about skill depreciation. Similarly, both of the earnings measures I use proxy real earnings in that they capture only a subset of the total variation in earnings. Finally, the period during which I record skill additions (2019 to May 2021) is short at two and a half years and includes the months when the Covid-19 pandemic impacted the US labor market. The latter feature might limit the external validity of the findings.

These findings open several paths for future work on how skill accumulation shapes careers. First, identification can be sharpened—e.g., leveraging firm training rollouts—to get some random variation in who gets to acquire skills. Second, richer measurement would be beneficial: longitudinal variation in standardized skill assessment data combined with rich career measures could help validate skill additions on LinkedIn as a measure of human-capital growth. This would also

allow dynamics such as depreciation to be traced over longer horizons. Finally, replicating the results in other countries, in post-pandemic cohorts, and during technology shocks (e.g., AI adoption) would enhance our understanding of where skill accumulation yields the largest mobility and earnings gains.

5.A Appendix A: Data and Variance Decomposition

5.A.1 Further Data

Education. LinkedIn profiles list degree type, field of study, and institution for each entry in the education section (see Figure 5.B.1). This data is used to derive three variables: highest degree, field of study for the highest degree, and college quality. Degrees are grouped into six levels in ascending order: high school, associate, bachelor, master, professional, and doctorate. We classify entries using online sources on degree abbreviations and hand-map common residual categories. For users with multiple degrees, we retain the highest ranked, using the most recent in case of ties. Fields of study are classified using the NCES CIP taxonomy, with manual mapping for unmatched entries. We assign college quality based on Times Higher Education's 2019 U.S. College Ranking.

Experience. From reported education and job spells, we calculate actual work experience as the time from graduation (associate or bachelor) to the 2019 scrape date during which a position was held. We also compute potential experience as the full time span since graduation.

Gender. Gender is predicted based solely from first names by Revelio Labs. More precisely, the prediction outputs probabilities that a worker is female. We classify a worker as female if this probability exceeds .5. Reassuringly, the distribution is bimodal with almost equal mass at probabilities 0 and 1.

Age. We estimate age based on the timing of the first educational degree. If a high school year is reported, we assume graduation at age 18. If not, we infer age 18 at the start of undergraduate studies or age 24 at the start of graduate studies, based on observed median lags between degrees.

Occupation. Revelio Labs aggregates the occupational information contained in every position on a profile to 336 occupation groups according to the 2018 SOC classification.

5.A.2 Decomposing the contribution of changes to positions to explained variance in earnings trajectories

We estimate the following regression model to quantify the association between job-switching patterns and year-on-year log earnings changes:

$$\Delta \ln Y_{it} = \sum_{g \in G} \beta_g \cdot D_{ig} + \alpha_t + \varepsilon_{it} \quad (5.3)$$

where:

- $\Delta \ln Y_{it}$ denotes the log change in earnings for worker i in year t ,
- D_{ig} is a dummy variable indicating whether worker i is in switch group $g \in G \setminus \{0\}$,
- the reference group $g = 0$ corresponds to “No change” in title, firm, or metro area,
- α_t are fixed effects for calendar-year transitions,
- ε_{it} is the residual.

Each worker belongs to exactly one switch group, so the predicted outcome \hat{y}_i for each individual is:

$$\hat{y}_i = \beta_g \quad \text{for the unique group } g \text{ such that } D_{ig} = 1$$

We are interested in the variance of these predicted values, which quantifies how much systematic wage change variation arises from differences in switch group. Because group membership is mutually exclusive, the distribution of predicted outcomes \hat{y}_i across workers is a discrete distribution with mass p_g (the share of individuals in group g) on value β_g . The variance of predicted values across the population is therefore:

$$\text{Var}(\hat{y}_i) = \sum_{g \in G} p_g \cdot (1 - p_g) \cdot \beta_g^2 \quad (5.4)$$

The first term represents the second moment of predicted values, and we use this to attribute group-level contributions. Specifically, we define the raw contribution of group g as:

$$\text{Contribution}_g = p_g \cdot (1 - p_g) \cdot \beta_g^2$$

and the relative contribution (or “share of explained variation”) as:

$$\text{Share of model variance}_g = \frac{p_g \cdot (1 - p_g) \cdot \beta_g^2}{\sum_{g \in G} p_g \cdot (1 - p_g) \cdot \beta_g^2}$$

This statistic reflects the fraction of the between-group second moment that can be attributed to group g .

5.B Appendix B: Robustness Checks

5.B.1 Intensive Margin Treatment

Table 5.B.1: Effect of an Additional Skill on Post-Treatment Earnings

Dependent variable	Earnings 2021	Earnings 2022	Earnings 2023
No. of skills added	58 (5)	71 (6)	63 (7)
R^2	0.744	0.678	0.644

Notes: $N = 5,141,520$. OLS regressions of predicted earnings on the number of skills added between 2019 and 2021. Specifications include the same controls and fixed effects as in Table 5.5: controls (gender, age, ratio of actual to potential experience, lagged earnings 2015–2019, baseline skill count and shares) and fixed effects (initial scrape month, highest degree, field of study, college ranking, metro area, start year of last job, occupation in 2019 and late 20s, graduation cohort, number of job switches). Robust standard errors in parentheses. Coefficients are dollars per additional skill.

Table 5.B.2: Effect of an Additional Skill on Post-Treatment ACS Earnings

Dependent variable	ACS Earnings 2021	ACS Earnings 2022	ACS Earnings 2023
No. of skills added	28 (14)	40 (20)	57 (25)
R^2	0.731	0.674	0.645

Notes: $N = 5,141,520$. OLS regressions of ACS occupation-level earnings on the number of skills added between 2019 and 2021. Specifications include the same controls and fixed effects as in Table 5.5: controls (gender, age, ratio of actual to potential experience, lagged ACS earnings 2015–2019, baseline skill count) and fixed effects (initial scrape month, highest degree, field of study, college ranking, metro area, start year of last job, occupation in 2019 and late 20s, graduation cohort, number of job switches). Standard errors in parentheses are clustered at the occupation level of the respective year. Coefficients are dollars per additional skill.

Table 5.B.3: OLS estimates for the effect of accumulating skills on job mobility (count specification)

Dependent variable (betw. 2019–2023)	# of position switches (1)	# of switches to higher-paying pos. (2)	# of occupation switches (3)	# of switches to higher-paying occ. (4)
No. of skills added	0.012 (0.0002)	0.007 (0.0001)	0.005 (0.0008)	0.003 (0.0004)
R^2	0.199	0.164	0.126	0.106

Notes: $N = 5,141,520$. OLS regressions of each outcome on the number of skills added (count) with the following controls and fixed effects. Fixed effects: five highest degrees obtained, 39 fields of study, five college ranking brackets, years of job tenure, 22 occupations, 154 MSAs, number of positions (all in 2019), month of initial scrape, 22 occupations at career start, graduation cohort. Controls: age in years, gender, ratio of actual to potential work experience, number of skills, skill shares (general, managerial, specific), lagged outcomes from 2015 to 2019. Standard errors in parentheses are robust in columns (1)–(2) and clustered at the 2021 occupation level for outcomes (3)–(4).

Table 5.B.4: Earnings and Skill Additions: Two-Part Specification

Dependent variable	Earnings 2021 (1)	Earnings 2022 (2)	Earnings 2023 (3)
τ (extensive margin)	663 (31)	904 (36)	938 (38)
δ (intensive margin)	-6 (5)	-16 (6)	-27 (6)
Adjusted R^2	0.744	0.678	0.644

Notes: $N = 5,141,520$. OLS regressions of predicted earnings on an indicator whether skills were added and an interaction of it with the number of skills added between 2019 and 2021 minus one. Specifications include the same controls and fixed effects as in Table 5.5: controls (gender, age, ratio of actual to potential experience, lagged earnings 2015–2019, baseline skill count and shares) and fixed effects (initial scrape month, highest degree, field of study, college ranking, metro area, start year of last job, occupation in 2019 and late 20s, graduation cohort, number of job switches). Robust standard errors in parentheses.

Table 5.B.5: ACS Earnings and Skill Additions: Two-Part Specification

Dependent variable	ACS Earnings 2021 (1)	ACS Earnings 2022 (2)	ACS Earnings 2023 (3)
τ (extensive margin)	340 (166)	574 (284)	649 (329)
δ (intensive margin)	-6 (14)	-14 (22)	-7 (25)
Adjusted R^2	0.731	0.674	0.645

Notes: $N = 5,141,520$. OLS regressions of ACS occupation-level earnings on an indicator whether skills were added and an interaction of it with the number of skills added between 2019 and 2021 minus one. Specifications include the same controls and fixed effects as in Table 5.5: controls (gender, age, ratio of actual to potential experience, lagged ACS earnings 2015–2019, baseline skill count) and fixed effects (initial scrape month, highest degree, field of study, college ranking, metro area, start year of last job, occupation in 2019 and late 20s, graduation cohort, number of job switches). Standard errors in parentheses are clustered at the occupation level of the respective year.

Table 5.B.6: OLS estimates for the extensive vs. intensive margin of skill accumulation on job mobility

Dependent variable (betw. 2019–2023)	# of position switches (1)	# of switches to higher-paying pos. (2)	# of occupation switches (3)	# of switches to higher-paying occ. (4)
τ (extensive margin)	0.068 (0.0010)	0.0436 (0.0008)	0.0195 (0.0066)	0.0131 (0.0036)
δ (intensive margin)	0.0047 (0.0002)	0.0026 (0.0001)	0.0028 (0.0005)	0.0015 (0.0002)
R^2	0.199	0.164	0.126	0.106

Notes: $N = 5,141,520$. Each column reports OLS estimates of the effect of skill accumulation on an indicator whether skills were added and an interaction of it with the number of skills added between 2019 and 2021 minus one. Fixed effects: five highest degrees obtained, 39 fields of study, five college ranking brackets, years of job tenure, 22 occupations, 154 MSAs, number of positions (all in 2019), month of initial scrape, 22 occupations at career start, graduation cohort. Controls: age in years, gender, ratio of actual to potential work experience, number of skills, skill shares (general, managerial, specific), lagged outcomes 2015–2019. Standard errors in parentheses are robust in (1)–(2) and clustered at the 2021 occupation level in (3)–(4).

Table 5.B.7: Earnings and Skill Additions: Mutually Exclusive Treatment Bins (ref. = 0)

Dependent variable	Earnings 2021 (1)	Earnings 2022 (2)	Earnings 2023 (3)
1 skill	651 (36)	877 (41)	906 (43)
2 skills	679 (71)	989 (82)	1,025 (87)
3–5 skills	631 (79)	838 (91)	866 (96)
6+ skills	633 (75)	749 (86)	630 (91)
Adjusted R^2	0.744	0.678	0.644

Notes: $N = 5,141,520$. OLS regressions of predicted earnings on bin dummies (reference = 0 skills). Specifications include the same controls and fixed effects as in Table 5.5: controls (gender, age, ratio of actual to potential experience, lagged earnings 2015–2019, baseline skill count and shares) and fixed effects (initial scrape month, highest degree, field of study, college ranking, metro area, start year of last job, occupation in 2019 and late 20s, graduation cohort, number of job switches). Robust standard errors in parentheses. Robust standard errors in parentheses.

5.B.2 Staleness

Table 5.B.8: Treatment Effect on Post-Treatment Earnings

Dependent variable	Earnings 2021	Earnings 2022	Earnings 2023
Skills added	739 (31)	1,012 (47)	1,058 (45)
Control mean	91,914	94,240	95,125
Treated - control mean	6,103	6,320	6,249
R ²	0.630	0.544	0.498

Notes: $N = 3,659,723$. OLS regressions of predicted earnings on a treatment indicator. Specifications include the same controls and fixed effects as in Table 5.5: controls (gender, age, ratio of actual to potential experience, lagged earnings 2015–2019, baseline skill count and shares) and fixed effects (initial scrape month, highest degree, field of study, college ranking, metro area, start year of last job, occupation in 2019 and late 20s, graduation cohort, number of job switches). Robust standard errors in parentheses.

Table 5.B.9: Treatment effect on Post-Treatment ACS Earnings

Dependent variable	Acs Earnings 2021	Acs Earnings 2022	Acs Earnings 2023
Skills added	424 (191)	673 (306)	823 (364)
Control mean	96,327	100,712	105,159
Treated - control mean	2,920	3,135	3,438
R ²	0.614	0.542	0.506

Notes: $N = 3,659,723$. OLS regressions of predicted earnings on a treatment indicator with controls and fixed effects. Fixed effects: initial scrape month, 5 highest degrees obtained, 39 fields of study, 5 college ranking brackets, years of job tenure, 22 occupations, 154 MSAs, number of positions since 2015 (all in 2019), 22 occupations at career start, graduation cohort. Controls: age in years, gender, ratio of actual to potential work experience, number of skills, skill shares for general, managerial, specific (all in 2019), lagged earnings from 2015 to 2019. Standard errors in parentheses are clustered at the occupation level of the respective year.

Table 5.B.10: Treatment Effect on job mobility outcomes

Dependent variable (betw. 2019–2023)	# of position switches (1)	# of switches to higher-paying pos. (2)	# of occupation switches (3)	# of switches to higher-paying occ. (4)
Skills added	0.068 (0.004)	0.044 (0.002)	0.017 (0.007)	0.013 (0.004)
Control mean	1.000	0.572	0.547	0.269
Treated - control mean	0.115	0.067	0.025	0.015
R ²	0.128	0.116	0.091	0.099

Notes: $N = 3,659,723$. OLS regressions of career outcomes on treatment indicator with the following controls and fixed effects. Fixed effects: five highest degrees obtained, 39 fields of study, five college ranking brackets, years of job tenure, 22 occupations, 154 MSAs, number of positions (all in 2019), month of initial scrape, 22 occupations at career start, graduation cohort. Controls: age in years, gender, ratio of actual to potential work experience, number of skills, skill shares (general, managerial, specific), lagged outcomes from 2015 to 2019. Standard errors in parentheses are robust in columns (1), (2) and clustered at the 2021 occupation level for outcomes (3), (4).

Table 5.B.11: Placebo Treatment Effects on Lagged Earnings (LinkedIn vs. ACS)

Dependent variable	Earnings 2019 (1)	AcS Earnings 2019 (2)
Skills added	83 (42)	107 (150)
Control mean	85,744	91,523
Treated - control mean	5,361	2,609
R^2	0.823	0.769

Notes: $N = 3,659,723$. OLS regressions of predicted/occupation level average earnings on a treatment indicator with the following controls and fixed effects. Fixed effects: 5 highest degrees obtained, 39 fields of study, 5 college ranking brackets, years of job tenure, 22 occupations, 154 MSAs, number of positions since 2015 (all in 2019), 22 occupations at career start, graduation cohort. Controls: age in years, gender, ratio of actual to potential work experience, number of skills, skill shares for general, managerial, specific (all in 2019), lagged earnings from 2015 to 2018. Robust (1)/clustered at the occupation level (2) standard errors in parentheses.

5.B.3 Covid

Table 5.B.12: Treatment Effect on Post-Treatment Earnings

Dependent variable	Earnings 2021	Earnings 2022	Earnings 2023
Skills Added	388 (43)	508 (62)	518 (60)
Control mean	90,689	92,695	93,699
Treated - control mean	-1,342	-827	-692
R^2	0.745	0.678	0.644

Notes: $N = 4,509,223$. OLS regressions of predicted earnings on a treatment indicator. Specifications include the same controls and fixed effects as in Table 5.5: controls (gender, age, ratio of actual to potential experience, lagged earnings 2015–2019, baseline skill count and shares) and fixed effects (initial scrape month, highest degree, field of study, college ranking, metro area, start year of last job, occupation in 2019 and late 20s, graduation cohort, number of job switches). Robust standard errors in parentheses.

Table 5.B.13: Treatment effect on Post-Treatment ACS Earnings

Dependent variable	AcS Earnings 2021	AcS Earnings 2022	AcS Earnings 2023
Skills Added	269 (113)	276 (163)	406 (198)
Control mean	95,813	100,326	104,905
Treated - control mean	86	-130	-32
R ²	0.735	0.678	0.648

Notes: $N = 4,509,223$. OLS regressions of predicted earnings on a treatment indicator with controls and fixed effects. Fixed effects: initial scrape month, 5 highest degrees obtained, 39 fields of study, 5 college ranking brackets, years of job tenure, 22 occupations, 154 MSAs, number of positions since 2015 (all in 2019), 22 occupations at career start, graduation cohort. Controls: age in years, gender, ratio of actual to potential work experience, number of skills, skill shares for general, managerial, specific (all in 2019), lagged earnings from 2015 to 2019. Standard errors in parentheses are clustered at the occupation level of the respective year.

Table 5.B.14: Treatment Effect on job mobility outcomes

Dependent variable (betw. 2019–2023)	# of position switches (1)	# of switches to higher-paying pos. (2)	# of occupation switches (3)	# of switches to higher-paying occ. (4)
Skills Added	0.102 (0.007)	0.064 (0.004)	0.042 (0.006)	0.025 (0.003)
Control mean	0.696	0.404	0.381	0.188
Treated - control mean	0.307	0.195	0.146	0.074
R ²	0.202	0.166	0.129	0.107

Notes: $N = 4,509,223$. OLS regressions of career outcomes on treatment indicator with the following controls and fixed effects. Fixed effects: five highest degrees obtained, 39 fields of study, five college ranking brackets, years of job tenure, 22 occupations, 154 MSAs, number of positions (all in 2019), month of initial scrape, 22 occupations at career start, graduation cohort. Controls: age in years, gender, ratio of actual to potential work experience, number of skills, skill shares (general, managerial, specific), lagged outcomes from 2015 to 2019. Standard errors in parentheses are robust in columns (1), (2) and clustered at the 2021 occupation level for outcomes (3), (4).

5.B.4 Reverse Causality

Table 5.B.15: Treatment Effect on Post-Treatment Earnings

Dependent variable	Earnings 2021	Earnings 2022	Earnings 2023
Skills Added	326 (47)	711 (52)	728 (43)
Control mean	90,689	92,695	93,699
Treated - control mean	6,650	7,110	7,058
R ²	0.754	0.686	0.650

Notes: $N = 5,054,348$. OLS regressions of predicted earnings on a treatment indicator. Specifications include the same controls and fixed effects as in Table 5.5: controls (gender, age, ratio of actual to potential experience, lagged earnings 2015–2019, baseline skill count and shares) and fixed effects (initial scrape month, highest degree, field of study, college ranking, metro area, start year of last job, occupation in 2019 and late 20s, graduation cohort, number of job switches). Robust standard errors in parentheses.

Table 5.B.16: Treatment effect on Post-Treatment ACS Earnings

Dependent variable	AcS Earnings 2021	AcS Earnings 2022	AcS Earnings 2023
Skills Added	311 (213)	487 (311)	548 (357)
Control mean	95,813	100,326	104,905
Treated - control mean	3,423	3,616	3,830
R^2	0.739	0.680	0.650

Notes: $N = 5,054,348$. OLS regressions of predicted earnings on a treatment indicator with controls and fixed effects. Fixed effects: initial scrape month, 5 highest degrees obtained, 39 fields of study, 5 college ranking brackets, years of job tenure, 22 occupations, 154 MSAs, number of positions since 2015 (all in 2019), 22 occupations at career start, graduation cohort. Controls: age in years, gender, ratio of actual to potential work experience, number of skills, skill shares for general, managerial, specific (all in 2019), lagged earnings from 2015 to 2019. Standard errors in parentheses are clustered at the occupation level of the respective year.

Table 5.B.17: Treatment Effect on job mobility outcomes

Dependent variable (betw. 2019–2023)	# of position switches (1)	# of switches to higher-paying pos. (2)	# of occupation switches (3)	# of switches to higher-paying occ. (4)
Skills Added	0.063 (0.005)	0.041 (0.003)	0.016 (0.006)	0.011 (0.004)
Control mean	0.696	0.404	0.381	0.188
Treated - control mean	0.164	0.098	0.059	0.031
R^2	0.201	0.166	0.127	0.106

Notes: $N = 5,054,348$. OLS regressions of career outcomes on treatment indicator with the following controls and fixed effects. Fixed effects: five highest degrees obtained, 39 fields of study, five college ranking brackets, years of job tenure, 22 occupations, 154 MSAs, number of positions (all in 2019), month of initial scrape, 22 occupations at career start, graduation cohort. Controls: age in years, gender, ratio of actual to potential work experience, number of skills, skill shares (general, managerial, specific), lagged outcomes from 2015 to 2019. Standard errors in parentheses are robust in columns (1), (2) and clustered at the 2021 occupation level for outcomes (3), (4).

Table 5.B.18: Placebo Treatment Effects on Lagged Earnings (LinkedIn vs. ACS)

Dependent variable	Earnings 2019 (1)	AcS Earnings 2019 (2)
Skills Added	117 (41)	66 (135)
Control mean	85,523	91,204
Treated - control mean	6,037	3,079
R^2	0.883	0.840

Notes: $N = 5,054,348$. OLS regressions of predicted/occupation level average earnings on a treatment indicator with the following controls and fixed effects. Fixed effects: 5 highest degrees obtained, 39 fields of study, 5 college ranking brackets, years of job tenure, 22 occupations, 154 MSAs, number of positions since 2015 (all in 2019), 22 occupations at career start, graduation cohort. Controls: age in years, gender, ratio of actual to potential work experience, number of skills, skill shares for general, managerial, specific (all in 2019), lagged earnings from 2015 to 2018. Robust (1)/clustered at the occupation level (2) standard errors in parentheses.

Skills

- System Architecture
- Enterprise Architecture
- Enterprise IT Infrastructure
- Teamwork
- scale
- Java
- Linux
- C++
- English
- Hadoop
- SOA
- Big Data
- High-performance computing
- SQL
- C#
- Algorithms
- German

Professional experience

- Google**
7 years 8 months
Munich, Bavaria, Germany
Software Engineer
Oct. 2019–Present · 5 years 10 months
- Strategic Cloud Engineer**
Dec. 2017–Oct. 2019 · 1 year 11 months
- DevOps Engineer**
Accenture
Aug. 2014–Nov. 2017 · 3 years 4 months
Munich, Bavaria, Germany
- Software Engineer**
Netlight
Feb. 2014–July 2014 · 6 months
Munich, Bavaria, Germany
- Data Scientist**
Allianz Global Investors
June 2013–Dec. 2013 · 7 months
Munich, Bavaria, Germany
- Technical Architect**
Siemens
Apr. 2012–Apr. 2013 · 1 year 1 month
Munich, Bavaria, Germany

Education

- Technische Universität München**
Master's Degree, Computational Science and Engineering
2011–2014
- Bavarian Graduate School of Computational Engineering**
Honours Degree, Computational Engineering
2012–2014
- Freie Universität Bozen**
Bachelor's Degree, Computer Science
2008–2011
- Franklin & Marshall College**
Exchange Semester, Computer Science
2010–2010

Figure 5.B.1: Profile
Notes: Appendix LinkedIn Profile

Overview of Tools Used

This thesis relies on multiple digital tools – mostly related to handling and analyzing large datasets as well as writing text.

Statistical analyses for this dissertation were conducted using the open-source programming language R on Windows 10 and 11. To enhance computational performance, computing resources at ifo Institute were used. In addition, Chapters 2 and 5 used Google Cloud services—Compute Engine, Cloud Storage, and BigQuery—for data wrangling and computation.

Writing and typesetting were managed in Overleaf. Large Language Models (specifically OpenAI’s ChatGPT) were used to refine and revise prose and to debug code and suggest implementation approaches. All code suggestions were reviewed, adapted as needed, and validated by the author; all analytical design choices, statistical results, and substantive conclusions are the author’s responsibility.

References

- (Destatis), S. B. (2022). "Berufliche Weiterbildung in Unternehmen - Sechste Europäische Erhebung über die berufliche Weiterbildung in Unternehmen (CVTS6) - 2020". In.
- Abadie, A. (2005). "Semiparametric Difference-in-Differences Estimators". In: *The Review of Economic Studies* 72.1, pp. 1–19. doi: 10.1111/0034-6527.00321.
- Abadie, A., S. Athey, G. W. Imbens, and J. M. Wooldridge (2022). "When Should You Adjust Standard Errors for Clustering?" In: *The Quarterly Journal of Economics* 138.1, pp. 1–35. doi: 10.1093/qje/qjac038.
- Abowd, J. M., F. Kramarz, and D. N. Margolis (1999). "High Wage Workers and High Wage Firms". In: *Econometrica* 67.2, pp. 251–333. JSTOR: 2999586.
- Acemoglu, D. and D. Autor (2011). "Skills, Tasks and Technologies: Implications for Employment and Earnings". In: *Handbook of Labor Economics*. Ed. by D. Card and O. Ashenfelter. Vol. 4. Elsevier, pp. 1043–1171. doi: 10.1016/S0169-7218(11)02410-5.
- Acemoglu, D. and J.-S. Pischke (1998). "Why Do Firms Train? Theory and Evidence". In: *The Quarterly Journal of Economics* 113.1, pp. 79–119. doi: 10.1162/003355398555531.
- (1999a). "The Structure of Wages and Investment in General Training". In: *Journal of Political Economy* 107.3, pp. 539–572. doi: 10.1086/250071. JSTOR: 10.1086/250071.
- Acemoglu, D. and J.-S. Pischke (1999b). "Beyond Becker: Training in Imperfect Labour Markets". In: *The Economic Journal* 109.453, pp. 112–142. doi: 10.1111/1468-0297.00405.
- Adda, J. and C. Dustmann (2023). "Sources of Wage Growth". In: *Journal of Political Economy* 131.2, pp. 456–503. doi: 10.1086/721657.
- Adda, J., C. Dustmann, and K. Stevens (2017). "The Career Costs of Children". In: *Journal of Political Economy* 125.2, pp. 293–337. doi: 10.1086/690952.
- Adhvaryu, A., N. Kala, and A. Nyshadham (2023). "Returns to On-the-Job Soft Skills Training". In: *Journal of Political Economy*. doi: 10.1086/724320.

- Ahn, T. and J. G. Trogon (2017). "Peer Delinquency and Student Achievement in Middle School". In: *Labour Economics* 44, pp. 192–217. doi: 10.1016/j.labeco.2017.01.006.
- Almlund, M., A. L. Duckworth, J. Heckman, and T. Kautz (2011). "Personality Psychology and Economics". In: *Handbook of the Economics of Education* 4, pp. 1–181. doi: 10.1016/B978-0-444-53444-6.00001-8.
- Altonji, J. G., P. Arcidiacono, and A. Maurel (2016). "Chapter 7 - The Analysis of Field Choice in College and Graduate School: Determinants and Wage Effects". In: *Handbook of the Economics of Education*. Ed. by E. A. Hanushek, S. Machin, and L. Woessmann. Vol. 5. Elsevier, pp. 305–396. doi: 10.1016/B978-0-444-63459-7.00007-5.
- Anderson, B. M. (2024). *Why It's Important to List Skills on Your LinkedIn Profile*. Retrieved April 17, 2025. URL: <https://www.linkedin.com/business/talent/blog/talent-acquisition/skills-on-linkedin-profile>.
- Angrist, J. D., P. A. Pathak, and C. R. Walters (2013). "Explaining Charter School Effectiveness". In: *American Economic Journal: Applied Economics* 5.4, pp. 1–27. doi: 10.1257/app.5.4.1.
- Angrist, J. D. and J.-S. Pischke (2009). *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton University Press.
- Arcidiacono, P., V. J. Hotz, A. Maurel, and T. Romano (2020). "Ex Ante Returns and Occupational Choice". In: *Journal of Political Economy* 128.12, pp. 4475–4522. doi: 10.1086/710559.
- Arellano-Bover, J. (2022). "The Effect of Labor Market Conditions at Entry on Workers' Long-Term Skills". In: *The Review of Economics and Statistics*, pp. 1–45. doi: 10.1162/rest_a_01008.
- (2024). "Career Consequences of Firm Heterogeneity for Young Workers: First Job and Firm Size". In: *Journal of Labor Economics* 42.2, pp. 549–589. doi: 10.1086/723500.
- Arellano-Bover, J. and F. Saltiel (2024). "Differences in On-the-Job Learning across Firms". In: *Journal of Labor Economics*, forthcoming. doi: 10.1086/732357.
- Arizona State University Alumni (2021). *The Five Most Important Things to Include in Your LinkedIn Profile*. Retrieved December 26, 2023. URL: <https://nursingandhealth.asu.edu/posts/alumni/five-most-important-things-include-your-linkedin-profile>.

- Arkhangelsky, D. and G. Imbens (2024). "Causal Models for Longitudinal and Panel Data: A Survey". In: *The Econometrics Journal* 27.3, pp. C1–C61. doi: 10.1093/ectj/utae014.
- Arnold, K.-H. and W. Vollstädt (2001). "Arbeits- Und Sozialverhalten in Der Schule. Möglichkeiten Und Grenzen Ihrer Beurteilung Durch "Kopfnoten"." In: *Die deutsche Schule* 93.2, pp. 199–209.
- Arntz, M., S. Blesse, and P. Doerrenberg (2023). "The End of Work Feels Near. How Do People Perceive the Impact of Digital Technologies and Automation?" In: *Working Paper*.
- Athey, S., J. Tibshirani, and S. Wager (2019). "Generalized Random Forests". In: *The Annals of Statistics* 47.2, pp. 1148–1178. doi: 10.1214/18-AOS1709.
- Attanasio, O. P. and K. M. Kaufmann (2014). "Education Choices and Returns to Schooling: Mothers' and Youths' Subjective Expectations and Their Role by Gender". In: *Journal of Development Economics* 109, pp. 203–216. doi: 10.1016/j.jdeveco.2014.04.003.
- Autor, D., C. Chin, A. Salomons, and B. Seegmiller (2024). "New Frontiers: The Origins and Content of New Work, 1940–2018*". In: *The Quarterly Journal of Economics* 139.3, pp. 1399–1465. doi: 10.1093/qje/qjae008.
- Autor, D., C. Goldin, and L. F. Katz (2020). "Extending the Race between Education and Technology". In: *AEA Papers and Proceedings* 110, pp. 347–351. doi: 10.1257/pandp.20201061.
- Autor, D. H. (2001). "Why Do Temporary Help Firms Provide Free General Skills Training?" In: *The Quarterly Journal of Economics* 116.4, pp. 1409–1448. doi: 10.1162/003355301753265615.
- (2014). "Skills, Education, and the Rise of Earnings Inequality among the "Other 99 Percent"". In: *Science* 344.6186, pp. 843–851. doi: 10.1126/science.1251868.
- Autor, D. H. and D. Dorn (2013). "The Growth of Low-Skill Service Jobs and the Polarization of the US Labor Market". In: *American Economic Review* 103.5, pp. 1553–1597. doi: 10.1257/aer.103.5.1553.
- Autor, D. H. and M. J. Handel (2013). "Putting Tasks to the Test: Human Capital, Job Tasks, and Wages". In: *Journal of Labor Economics* 31.S1, S59–S96. doi: 10.1086/669332.
- Autor, D. H., F. Levy, and R. J. Murnane (2003). "The Skill Content of Recent Technological Change: An Empirical Exploration". In: *The Quarterly Journal of Economics* 118.4, pp. 1279–1333. doi: 10.1162/003355303322552801.

- Auxier, B. and M. Anderson (2021). *Social Media Use in 2021*. Retrieved July 15, 2024. Pew Research Center. URL: <https://www.pewresearch.org/internet/2021/04/07/social-media-use-in-2021/>.
- Azmat, G., V. Cuñat, and E. Henry (2025). "Gender Promotion Gaps and Career Aspirations". In: *Management Science* 71.3, pp. 2127–2141. DOI: 10.1287/mnsc.2023.00715.
- Azmat, G. and R. Ferrer (2017). "Gender Gaps in Performance: Evidence from Young Lawyers". In: *Journal of Political Economy* 125.5, pp. 1306–1355. DOI: 10.1086/693686.
- BA (2024). *Statistik Der Bundesagentur Für Arbeit. Beschäftigungsstatistik, Monatsbericht: Mai 2024*. Tech. rep. BA (German Federal Employment Agency).
- Baier, H. A., P. Lergertporer, and T. Rittmannsberger (2025). "Firms' Expectations about Skill Shortages". In: *Small Business Economics*, forthcoming.
- Baker, G., M. Gibbs, and B. Holmstrom (1994). "The Internal Economics of the Firm: Evidence from Personnel Data". In: *The Quarterly Journal of Economics* 109.4, pp. 881–919. DOI: 10.2307/2118351.
- Baker, R., E. Bettinger, B. Jacob, and I. Marinescu (2018). "The Effect of Labor Market Information on Community College Students' Major Choice". In: *Economics of Education Review* 65, pp. 18–30. DOI: 10.1016/j.econedurev.2018.05.005.
- Bandiera, O., N. Parekh, B. Petrongolo, and M. Rao (2022). "Men Are from Mars, and Women Too: A Bayesian Meta-analysis of Overconfidence Experiments". In: *Economica* 89.S1, S38–S70. DOI: 10.1111/ecca.12407.
- Barron, J. M., D. A. Black, and M. A. Loewenstein (1993). "Gender Differences in Training, Capital, and Wages". In: *The Journal of Human Resources* 28.2, pp. 343–364. DOI: 10.2307/146207. JSTOR: 146207.
- Bartel, A. P. (1995). "Training, Wage Growth, and Job Performance: Evidence from a Company Database". In: *Journal of Labor Economics* 13.3, pp. 401–425. JSTOR: 2535150.
- Bassanini, A., A. L. Booth, G. Brunello, M. De Paola, and E. Leuven (2007). "Workplace Training in Europe". In: *Education and Training in Europe*, ed. G. Brunello, P. Garibaldi, and E. Wasmer, pp. 143–323. DOI: 10.2139/ssrn.756405.
- Battisti, M., C. Dustmann, and U. Schönberg (2023). "Technological and Organizational Change and the Careers of Workers". In: *Journal of the European Economic Association* forthcoming.

- Baumert, J. et al., eds. (2002). *PISA 2000 — Die Länder der Bundesrepublik Deutschland im Vergleich*. Wiesbaden: VS Verlag für Sozialwissenschaften. DOI: 10.1007/978-3-663-11042-2.
- Bayer, C. and M. Kuhn (2023). "Job Levels and Wages". In: *SSRN Electronic Journal*. DOI: 10.2139/ssrn.4464590.
- Becker, A., T. Deckers, T. Dohmen, A. Falk, and F. Kosse (2012). "The Relationship Between Economic Preferences and Psychological Personality Measures". In: *Annual Review of Economics* 4.1, pp. 453–478. DOI: 10.1146/annurev-economics-080511-110922.
- Becker, G. S. (1962). "Investment in Human Capital: A Theoretical Analysis". In: *Journal of Political Economy* 70.5, pp. 9–49. JSTOR: 1829103.
- (1964). *Human Capital: A Theoretical and Empirical Analysis, with Special Reference to Education*. Ed. by 3. edition. Chicago, IL: University of Chicago Press.
- Becker, G. S. and B. R. Chiswick (1966). "Education and the Distribution of Earnings". In: *The American Economic Review* 56.1/2, pp. 358–369. JSTOR: 1821299.
- Belot, M., P. Kircher, and P. Muller (2019). "Providing Advice to Jobseekers at Low Cost: An Experimental Study on Online Advice". In: *The Review of Economic Studies* 86.4, pp. 1411–1447. DOI: 10.1093/restud/rdy059.
- Ben-Porath, Y. (1967). "The Production of Human Capital and the Life Cycle of Earnings". In: *Journal of Political Economy* 75.4, Part 1, pp. 352–365. DOI: 10.1086/259291.
- Benson, A., D. Li, and K. Shue (2019). "Promotions and the Peter Principle". In: *The Quarterly Journal of Economics* 134.4, pp. 2085–2134. DOI: 10.1093/qje/qjz022.
- Bertrand, M. (2018). "Coase Lecture – The Glass Ceiling". In: *Economica* 85.338, pp. 205–231. DOI: 10.1111/ecca.12264.
- Bertrand, M., C. Goldin, and L. F. Katz (2010). "Dynamics of the Gender Gap for Young Professionals in the Financial and Corporate Sectors". In: *American Economic Journal: Applied Economics* 2.3, pp. 228–255. DOI: 10.1257/app.2.3.228.
- Best Universities (2024). *Abbreviations for Degrees and Other Academic Distinctions*. Retrieved January 8, 2024. URL: <https://best-universities.net/degrees/degree-abbreviations/>.
- Biewen, M., B. Fitzenberger, A. Osikominu, and M. Paul (2014). "The Effectiveness of Public-Sponsored Training Revisited: The Importance of Data and Methodological Choices". In: *Journal of Labor Economics* 32.4, pp. 837–897. DOI: 10.1086/677233. JSTOR: 10.1086/677233.

- Bird, K. A. et al. (2021). "Nudging at Scale: Experimental Evidence from FAFSA Completion Campaigns". In: *Journal of Economic Behavior & Organization* 183, pp. 105–128. doi: 10.1016/j.jebo.2020.12.022.
- Black, D. A., L. Skipper, and J. A. Smith (2023). "Firm Training". In: *Handbook of the Economics of Education*. Ed. by E. A. Hanushek, S. Machin, and L. Woessmann. Vol. 7. Elsevier, pp. 287–468. doi: 10.1016/bs.hesedu.2023.03.004.
- Blandin, A. (2018). "Learning by Doing and Ben-Porath: Life-cycle Predictions and Policy Implications". In: *Journal of Economic Dynamics and Control* 90, pp. 220–235. doi: 10.1016/j.jedc.2018.03.002.
- Blass, A. A., S. Lach, and C. F. Manski (2010). "Using Elicited Choice Probabilities to Estimate Random Utility Models: Preferences for Electricity Reliability*". In: *International Economic Review* 51.2, pp. 421–440. doi: 10.1111/j.1468-2354.2010.00586.x.
- Blau, F. D. and L. M. Kahn (2017). "The Gender Wage Gap: Extent, Trends, and Explanations". In: *Journal of Economic Literature*, p. 77.
- Blossfeld, H.-P. and H.-G. Roßbach, eds. (2019). *Education as a Lifelong Process: The German National Educational Panel Study (NEPS)*. Vol. 3. Edition ZfE. Wiesbaden: Springer Fachmedien. doi: 10.1007/978-3-658-23162-0.
- BMBF (2024). "Bundesministerium für Bildung und Forschung: Weiterbildungsverhalten in Deutschland 2022 - Ergebnisse des Adult Education Survey — AES-Trendbericht". In.
- Böckerman, P. and P. Ilmakunnas (2009). "Job Disamenities, Job Satisfaction, Quit Intentions, and Actual Separations: Putting the Pieces Together". In: *Industrial Relations: A Journal of Economy and Society* 48.1, pp. 73–96. doi: 10.1111/j.1468-232X.2008.00546.x.
- Boneva, T., M. Golin, and C. Rauh (2022). "Can Perceived Returns Explain Enrollment Gaps in Postgraduate Education?" In: *Labour Economics*. European Association of Labour Economists, World Conference EALE/SOLE/AASLE, Berlin, Germany, 25 – 27 June 2020 77, p. 101998. doi: 10.1016/j.labeco.2021.101998.
- Boneva, T. and C. Rauh (2021). "Socio-Economic Gaps in University Enrollment: The Role of Perceived Pecuniary and Non-Pecuniary Returns". In: *Working Paper*, p. 77.
- Bordalo, P., K. Coffman, N. Gennaioli, and A. Shleifer (2019). "Beliefs about Gender". In: *American Economic Review* 109.3, pp. 739–773. doi: 10.1257/aer.20170007.

- Borghans, L., B. H. H. Golsteyn, J. J. Heckman, and J. E. Humphries (2016). "What Grades and Achievement Tests Measure". In: *Proceedings of the National Academy of Sciences* 113.47, pp. 13354–13359. doi: 10.1073/pnas.1601135113.
- Bowles, S. and H. Gintis (2002). "Schooling in Capitalist America Revisited". In: *Sociology of Education* 75.1, pp. 1–18. doi: 10.2307/3090251. JSTOR: 3090251.
- Bowles, S., H. Gintis, and M. Osborne (2001). "The Determinants of Earnings: A Behavioral Approach". In: *Journal of Economic Literature* 39.4, pp. 1137–1176. doi: 10.1257/jel.39.4.1137.
- Braxton, J. C. and B. Taska (2023). "Technological Change and the Consequences of Job Loss". In: *American Economic Review* 113.2, pp. 279–316. doi: 10.1257/aer.20210182.
- Breiman, L. (2001). "Random Forests". In: *Machine Learning* 45.1, pp. 5–32.
- Bronson, M. A. and P. S. Thoursie (2019). "The Wage Growth and Within-Firm Mobility of Men and Women: New Evidence and Theory". In: *Journal of Human Resources* 3.4, pp. 1–24. doi: 10.3386/w33396. eprint: 33396.
- Bureau of Labor Statistics (2022). *Educational Attainment for Workers 25 Years and Older by Detailed Occupation*. [retrieved July 15, 2024]. URL: <https://www.bls.gov/emp/tables/educational-attainment.htm>.
- Caldwell, S., I. Haegele, and J. Heining (2025a). *Bargaining and Inequality in the Labor Market*. Working Paper. doi: 10.3386/w33396. eprint: 33396.
- (2025b). *Firm Pay and Worker Search*. Working Paper. doi: 10.3386/w33445. eprint: 33445.
- Caliendo, M., D. A. Cobb-Clark, C. Obst, H. Seitz, and A. Uhlendorff (2020). "Locus of Control and Investment in Training". In: *Journal of Human Resources*, 0318–9377R2. doi: 10.3368/jhr.57.4.0318-9377R2.
- Caliendo, M., D. A. Cobb-Clark, C. Obst, and A. Uhlendorff (2023). "Risk Preferences and Training Investments". In: *Journal of Economic Behavior & Organization* 205, pp. 668–686. doi: 10.1016/j.jebo.2022.11.024.
- Caliendo, M., D. A. Cobb-Clark, H. Pfeifer, A. Uhlendorff, and C. Wehner (2024). "Managers' Risk Preferences and Firm Training Investments". In: *European Economic Review* 161, p. 104616. doi: 10.1016/j.euroecorev.2023.104616.
- Caliendo, M. et al. (2025). "When Managers Choose: Gender Disparities in Employer Training Provision". In: *CEPA Discussion Papers; 90*, 4334 KB, 80 pages. doi: 10.25932/PUBLISHUP-68304.
- Callaway, B. and P. H. C. Sant'Anna (2021). "Difference-in-Differences with Multiple Time Periods". In: *Journal of Econometrics*. Themed Issue: Treatment Effect 1 225.2, pp. 200–230. doi: 10.1016/j.jeconom.2020.12.001.

- Cameron, A. C., J. B. Gelbach, and D. L. Miller (2008). "Bootstrap-Based Improvements for Inference with Clustered Errors". In: *The Review of Economics and Statistics* 90.3, pp. 414–427. doi: 10.1162/rest.90.3.414.
- Card, D. (1999). "The Causal Effect of Education on Earnings". In: *Handbook of Labor Economics*. Vol. 3. Elsevier, pp. 1801–1863. doi: 10.1016/S1573-4463(99)03011-4.
- Card, D. and A. B. Krueger (1994). "Minimum Wages and Employment: A Case Study of the Fast-Food Industry in New Jersey and Pennsylvania". In: *The American Economic Review* 84.4, pp. 772–793. JSTOR: 2118030.
- Card, D., A. Mas, E. Moretti, and E. Saez (2012). "Inequality at Work: The Effect of Peer Salaries on Job Satisfaction". In: *American Economic Review* 102.6, pp. 2981–3003. doi: 10.1257/aer.102.6.2981.
- Castex, G. and E. Kogan Dechter (2014). "The Changing Roles of Education and Ability in Wage Determination". In: *Journal of Labor Economics* 32.4, pp. 685–710. doi: 10.1086/676018. JSTOR: 10.1086/676018.
- Chetty, R., D. J. Deming, and J. N. Friedman (2023). *Diversifying Society's Leaders? The Determinants and Causal Effects of Admission to Highly Selective Private Colleges*. Working Paper. doi: 10.3386/w31492. eprint: 31492.
- Cheung, C.-k. and S.-c. Llu (2000). "Acculturation, Social Integration and School Achievement among Low-ability Seventh Graders' School Achievement in Hong Kong". In: *International Journal of Adolescence and Youth* 8.1, pp. 81–108. doi: 10.1080/02673843.2000.9747843.
- Christoph, B., B. Matthes, and C. Ebner (2020). "Occupation-Based Measures—An Overview and Discussion". In: *KZfSS Kölner Zeitschrift für Soziologie und Sozialpsychologie* 72.1, pp. 41–78. doi: 10.1007/s11577-020-00673-4.
- Cinelli, C. and C. Hazlett (2020). "Making Sense of Sensitivity: Extending Omitted Variable Bias". In: *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 82.1, pp. 39–67. doi: 10.1111/rssb.12348.
- Close, D. (2009). "Fair Grades". In: *Teaching Philosophy* 32.4, pp. 361–398. doi: 10.5840/teachphil200932439.
- Cohen, J., A. C. Johnston, and A. Lindner (2025). "Skill Depreciation during Unemployment: Evidence from Panel Data". In: *American Economic Journal: Applied Economics* 17.3, pp. 208–235. doi: 10.1257/app.20230195.
- Conlon, J. J., L. Pilossoph, M. Wiswall, and B. Zafar (2018). *Labor Market Search With Imperfect Information and Learning*. Tech. rep. w24988. National Bureau of Economic Research. doi: 10.3386/w24988.

- Conzelmann, J. G. et al. (2022). *Grads on the Go: Measuring College-Specific Labor Markets for Graduates*. Working Paper 30088. National Bureau of Economic Research. doi: 10.3386/w30088.
- Cook, C., R. Diamond, J. V. Hall, J. A. List, and P. Oyer (2021). "The Gender Earnings Gap in the Gig Economy: Evidence from over a Million Rideshare Drivers". In: *The Review of Economic Studies* 88.5, pp. 2210–2238. doi: 10.1093/restud/rdaa081.
- Cortés, P. and J. Pan (2023). "Children and the Remaining Gender Gaps in the Labor Market". In: *Journal of Economic Literature* 61.4, pp. 1359–1409. doi: 10.1257/jel.20221549.
- Crump, R. K., V. J. Hotz, G. W. Imbens, and O. A. Mitnik (2009). "Dealing with Limited Overlap in Estimation of Average Treatment Effects". In: *Biometrika* 96.1, pp. 187–199. doi: 10.1093/biomet/asn055.
- Cullen, Z. and R. Perez-Truglia (2022). "How Much Does Your Boss Make? The Effects of Salary Comparisons". In: *Journal of Political Economy* 130.3, pp. 766–822. doi: 10.1086/717891.
- Cunha, F. and J. Heckman (2007). "The Technology of Skill Formation". In: *American Economic Review* 97.2, pp. 31–47. doi: 10.1257/aer.97.2.31.
- Currie, J. M. (2004). *Welfare and the Well-Being of Children*. London: Routledge. doi: 10.4324/9780203987575.
- Dale, S. B. and A. B. Krueger (2002). "Estimating the Payoff to Attending a More Selective College: An Application of Selection on Observables and Unobservables*". In: *The Quarterly Journal of Economics* 117.4, pp. 1491–1527. doi: 10.1162/003355302320935089.
- de Chaisemartin, C. and X. D'Haultfœuille (2020). "Two-Way Fixed Effects Estimators with Heterogeneous Treatment Effects". In: *American Economic Review* 110.9, pp. 2964–2996. doi: 10.1257/aer.20181169.
- De Fraja, G., T. Oliveira, and L. Zanchi (2010). "Must Try Harder: Evaluating the Role of Effort in Educational Attainment". In: *The Review of Economics and Statistics* 92.3, pp. 577–597. doi: 10.1162/REST_a_00013.
- de Quidt, J., J. Haushofer, and C. Roth (2018). "Measuring and Bounding Experiment Demand". In: *American Economic Review* 108.11, pp. 3266–3302. doi: 10.1257/aer.20171330.
- Dearden, L., H. Reed, and J. Van Reenen (2006). "The Impact of Training on Productivity and Wages: Evidence from British Panel Data". In: *Oxford Bulletin of Economics and Statistics* 68.4, pp. 397–421. doi: 10.1111/j.1468-0084.2006.00170.x.

- Degraux, X. (2023). *Recruitment, Job Search. . . : Do You Have the Skills?* Retrieved April 17, 2025. URL: <https://www.linkedin.com/pulse/recruitment-job-search-do-you-have-skills-xavier-degraux/>.
- Deming, D. and L. B. Kahn (2018). "Skill Requirements across Firms and Labor Markets: Evidence from Job Postings for Professionals". In: *Journal of Labor Economics* 36.S1, S337–S369. DOI: 10.1086/694106.
- Deming, D. J. (2024). "Why Do Wages Grow Faster for Educated Workers?" In: *Journal of Labor Economics, forthcoming*.
- Deming, D. J. and K. Noray (2020). "Earnings Dynamics, Changing Job Skills, and STEM Careers". In: *The Quarterly Journal of Economics* 135.4, pp. 1965–2005. DOI: 10.1093/qje/qjaa021.
- Deming, D. J. (2017). "The Growing Importance of Social Skills in the Labor Market". In: *The Quarterly Journal of Economics* 132.4, pp. 1593–1640. DOI: 10.1093/qje/qjx022.
- (2022). "Four Facts about Human Capital". In: *Journal of Economic Perspectives* 36.3, pp. 75–102. DOI: 10.1257/jep.36.3.75.
- (2023a). "Chapter 6 - Multidimensional Human Capital and the Wage Structure". In: *Handbook of the Economics of Education*. Ed. by E. A. Hanushek, S. Machin, and L. Woessmann. Vol. 7. Elsevier, pp. 469–504. DOI: 10.1016/bs.hesedu.2023.03.005.
- (2023b). "Multidimensional Human Capital and the Wage Structure". In: *Handbook of the Economics of Education*. Vol. 7. DOI: 10.3386/w31001.
- Deming, D. J. and M. I. Silliman (2024). *Skills and Human Capital in the Labor Market*. Working Paper. DOI: 10.3386/w32908. eprint: 32908.
- Diaz, B. S., A. Neyra Nazarrett, J. Ramirez, R. Sadun, and J. A. Tamayo (2025). *Training within Firms*. Working Paper. DOI: 10.3386/w33670. eprint: 33670.
- Dinerstein, M., R. Megalokonomou, and C. Yannelis (2022). "Human Capital Depreciation and Returns to Experience". In: *American Economic Review* 112.11, pp. 3725–3762. DOI: 10.1257/aer.20201571.
- Dobbie, W. and R. G. Fryer (2020). "Charter Schools and Labor Market Outcomes". In: *Journal of Labor Economics* 38.4, pp. 915–957. DOI: 10.1086/706534.
- Dohmen, T. and T. Jagelka (2024). "Accounting for Individual-Specific Reliability of Self-Assessed Measures of Economic Preferences and Personality Traits". In: *Journal of Political Economy Microeconomics* 2.3, pp. 399–462. DOI: 10.1086/727559.

- Dominitz, J. and C. F. Manski (1996). "Eliciting Student Expectations of the Returns to Schooling". In: *The Journal of Human Resources* 31.1, pp. 1–26. doi: 10.2307/146041. JSTOR: 146041.
- Dorn, D., F. Schoner, M. Seebacher, L. Simon, and L. Woessmann (2025). "Multi-dimensional Skills on LinkedIn Profiles: Measuring Human Capital and the Gender Skill Gap". In: *CESifo Working Paper No. 11846*.
- Duckworth, A. L., P. D. Quinn, and E. Tsukayama (2012). "What No Child Left Behind Leaves Behind: The Roles of IQ and Self-Control in Predicting Standardized Achievement Test Scores and Report Card Grades". In: *Journal of educational psychology* 104.2, pp. 439–451. doi: 10.1037/a0026280.
- Edin, P.-A., P. Fredriksson, M. Nybom, and B. Öckert (2022). "The Rising Return to Noncognitive Skill". In: *American Economic Journal: Applied Economics* 14.2, pp. 78–100. doi: 10.1257/app.20190199.
- Edin, P.-A. and M. Gustavsson (2008). "Time out of Work and Skill Depreciation". In: *Industrial and Labor Relations Review* 61.2, pp. 163–180. JSTOR: 25249132.
- Eggenberger, C., M. Rinawi, and U. Backes-Gellner (2018). "Occupational Specificity: A New Measurement Based on Training Curricula and Its Effect on Labor Market Outcomes". In: *Labour Economics* 51, pp. 97–107. doi: 10.1016/j.labeco.2017.11.010.
- Enke, B. and T. Graeber (2023). "Cognitive Uncertainty". In: *The Quarterly Journal of Economics* 138.4, pp. 2021–2067. doi: 10.1093/qje/qjad025.
- Enquete-Kommission „Zukunft des Bürgerschaftlichen Engagements“ des Deutschen Bundestages (2002). "Bürgerschaftliches Engagement: auf dem Weg in eine zukunftsfähige Bürgergesellschaft". In: *Bericht. Bürgerschaftliches Engagement: auf dem Weg in eine zukunftsfähige Bürgergesellschaft*. Wiesbaden: VS Verlag für Sozialwissenschaften, pp. 55–154. doi: 10.1007/978-3-322-92328-8_1.
- Espinosa, M. and C. T. Stanton (2022). *Training, Communications Patterns, and Spillovers Inside Organizations*. Working Paper. doi: 10.3386/w30224. eprint: 30224.
- Evsyukova, Y., F. Rusche, and W. Mill (2025). "LinkedOut? A Field Experiment on Discrimination in Job Network Formation*". In: *The Quarterly Journal of Economics* 140.1, pp. 283–334. doi: 10.1093/qje/qjae035.
- Exley, C. L. and J. B. Kessler (2022). "The Gender Gap in Self-Promotion*". In: *The Quarterly Journal of Economics*, qjac003. doi: 10.1093/qje/qjac003.
- Facchinello, L. (2020). *Short- and Long-run Effects of Early Grades*. SSRN Scholarly Paper ID 2966571. Rochester, NY: Social Science Research Network. doi: 10.2139/ssrn.2966571.

- Falk, A. and F. Zimmermann (2017). "Consistency as a Signal of Skills". In: *Management Science* 63.7, pp. 2197–2210. doi: 10.1287/mnsc.2016.2459.
- Feng, A. and G. Graetz (2017). "A Question of Degree: The Effects of Degree Class on Labor Market Outcomes". In: *Economics of Education Review* 61, pp. 140–161. doi: 10.1016/j.econedurev.2017.07.003.
- Ferman, B. and L. F. Fontes (2022). "Assessing Knowledge or Classroom Behavior? Evidence of Teachers' Grading Bias". In: *Journal of Public Economics* 216, p. 104773. doi: 10.1016/j.jpubeco.2022.104773.
- Freier, R., M. Schumann, and T. Siedler (2015). "The Earnings Returns to Graduating with Honors — Evidence from Law Graduates". In: *Labour Economics*. European Association of Labour Economists 26th Annual Conference 34, pp. 39–50. doi: 10.1016/j.labeco.2015.03.001.
- Fryer, R. G. (2011). "Financial Incentives and Student Achievement: Evidence from Randomized Trials*". In: *The Quarterly Journal of Economics* 126.4, pp. 1755–1798. doi: 10.1093/qje/qjr045.
- Gathmann, C. and U. Schönberg (2010). "How General Is Human Capital? A Task-Based Approach". In: *Journal of Labor Economics* 28.1, pp. 1–49. doi: 10.1086/649786.
- Gibbons, R. and M. Waldman (1999). "A Theory of Wage and Promotion Dynamics Inside Firms*". In: *The Quarterly Journal of Economics* 114.4, pp. 1321–1358. doi: 10.1162/003355399556287.
- (2006). "Enriching a Theory of Wage and Promotion Dynamics inside Firms". In: *Journal of Labor Economics* 24.1, pp. 59–107. doi: 10.1086/497819. JSTOR: 10.1086/497819.
- Giustinelli, P. (2016). "Group Decision Making with Uncertain Outcomes: Unpacking Child–Parent Choice of the High School Track". In: *International Economic Review* 57.2, pp. 573–602. doi: 10.1111/iere.12168.
- (2023). "Expectations in Education". In: *Handbook of Economic Expectations*. Ed. by R. Bachmann, G. Topa, and W. van der Klaauw. Academic Press, pp. 193–224. doi: 10.1016/B978-0-12-822927-9.00014-8.
- Gneezy, U. et al. (2019). "Measuring Success in Education: The Role of Effort on the Test Itself". In: *American Economic Review: Insights* 1.3, pp. 291–308. doi: 10.1257/aeri.20180633.
- Goebel, J. et al. (2019). "The German Socio-Economic Panel (SOEP)". In: *Jahrbücher für Nationalökonomie und Statistik* 239.2, pp. 345–360. doi: 10.1515/jbnst-2018-0022.

- Goldin, C. (2014). "A Grand Gender Convergence: Its Last Chapter". In: *American Economic Review* 104.4, pp. 1091–1119. doi: 10.1257/aer.104.4.1091.
- (2024). "Nobel Lecture: An Evolving Economic Force". In: *American Economic Review* 114.6, pp. 1515–1539. doi: 10.1257/aer.114.6.1515.
- Goldin, C. and L. Katz (2007). *The Race between Education and Technology: The Evolution of U.S. Educational Wage Differentials, 1890 to 2005*. Tech. rep. w12984. Cambridge, MA: National Bureau of Economic Research, w12984. doi: 10.3386/w12984.
- Goldin, C., S. P. Kerr, C. Olivetti, and E. Barth (2017). "The Expanding Gender Earnings Gap: Evidence from the LEHD-2000 Census". In: *American Economic Review* 107.5, pp. 110–114. doi: 10.1257/aer.p20171065.
- Golin, M. and C. Rauh (2022). "The Impact of Fear of Automation". In: *Working Paper*.
- Goodman-Bacon, A. (2021). "Difference-in-Differences with Variation in Treatment Timing". In: *Journal of Econometrics*. Themed Issue: Treatment Effect 1 225.2, pp. 254–277. doi: 10.1016/j.jeconom.2021.03.014.
- Goos, M., A. Manning, and A. Salomons (2014). "Explaining Job Polarization: Routine-Biased Technological Change and Offshoring". In: *American Economic Review* 104.8, pp. 2509–26. doi: 10.1257/aer.104.8.2509.
- Gregory, V. (2020). *Firms as Learning Environments: Implications for Earnings Dynamics and Job Search*. SSRN Scholarly Paper. Rochester, NY. doi: 10.20955/wp.2020.036. eprint: 3718086.
- Grewenig, E., P. Lergtporer, L. Simon, K. Werner, and L. Woessmann (2023). "Can Internet Surveys Represent the Entire Population? A Practitioners' Analysis". In: *European Journal of Political Economy* 78, p. 102382. doi: 10.1016/j.ejpoléco.2023.102382.
- Groes, F., P. Kircher, and I. Manovskii (2015). "The U-Shapes of Occupational Mobility". In: *The Review of Economic Studies* 82.2, pp. 659–692. doi: 10.1093/restud/rdu037.
- Gulyas, A. and K. Pytka (2025). "Understanding the Heterogeneity of Earnings Losses After Job Displacement: A Machine-Learning Approach". In: *Working Paper*.
- Guo, Y. M., O. Falck, C. Langer, V. Lindlacher, and S. Wiederhold (2024). "Training, Automation, and Wages: Worker-Level Evidence". In: *Working Paper*.
- Güvenen, F., F. Karahan, S. Ozkan, and J. Song (2021). "What Do Data on Millions of U.S. Workers Reveal About Lifecycle Earnings Dynamics?" In: *Econometrica* 89.5, pp. 2303–2339. doi: 10.3982/ECTA14603.

- Guvenen, F., B. Kuruscu, S. Tanaka, and D. Wiczer (2020). "Multidimensional Skill Mismatch". In: *American Economic Journal: Macroeconomics* 12.1, pp. 210–244. doi: 10.1257/mac.20160241.
- Haaland, I., C. Roth, and J. Wohlfart (2023). "Designing Information Provision Experiments". In: *Journal of Economic Literature* 61.1, pp. 3–40. doi: 10.1257/jel.20211658.
- Hall, R. E. and A. I. Mueller (2018). "Wage Dispersion and Search Behavior: The Importance of Nonwage Job Values". In: *Journal of Political Economy* 126.4, pp. 1594–1637. doi: 10.1086/697739.
- Hansen, A. T., U. Hvidman, and H. H. Sievertsen (2023). "Grades and Employer Learning". In: *Journal of Labor Economics* forthcoming. doi: 10.1086/724048.
- Hanushek, E. A. (1986). "The Economics of Schooling: Production and Efficiency in Public Schools". In: *Journal of Economic Literature* 24.3, pp. 1141–1177. JSTOR: 2725865.
- (2020). "Education Production Functions". In: *The Economics of Education (Second Edition) - A Comprehensive Overview*. Ed. by S. Bradley and C. Green, pp. 161–170. doi: 10.1016/B978-0-12-815391-8.00013-6.
- Hanushek, E. A., L. Kinne, F. Witthöft, and L. Woessmann (2025). "Age and Cognitive Skills: Use It or Lose It". In: *Science Advances* 11.10, eads1560. doi: 10.1126/sciadv.ads1560.
- Hanushek, E. A., G. Schwerdt, S. Wiederhold, and L. Woessmann (2015). "Returns to Skills around the World: Evidence from PIAAC". In: *European Economic Review* 73, pp. 103–130. doi: 10.1016/j.euroecorev.2014.10.006.
- Hanushek, E. A., G. Schwerdt, L. Woessmann, and L. Zhang (2017). "General Education, Vocational Education, and Labor-Market Outcomes over the Lifecycle". In: *Journal of Human Resources* 52.1, pp. 48–87. doi: 10.3368/jhr.52.1.0415-7074R.
- Hanushek, E. A. and L. Woessmann (2008). "The Role of Cognitive Skills in Economic Development". In: *Journal of Economic Literature* 46.3, pp. 607–668. JSTOR: 27647039.
- (2012). "Do Better Schools Lead to More Growth? Cognitive Skills, Economic Outcomes, and Causation". In: *Journal of Economic Growth* 17.4, pp. 267–321. doi: 10.1007/s10887-012-9081-x.
- Hastie, T., R. Tibshirani, and J. Friedman (2009). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Second Edition. New York, NY: Springer.

- Heckman, J. J., J. Stixrud, and S. Urzua (2006a). "The Effects of Cognitive and Noncognitive Abilities on Labor Market Outcomes and Social Behavior". In: *Journal of Labor Economics* 24.3, pp. 411–482.
- Heckman, J. J., H. Ichimura, and P. E. Todd (1997). "Matching As An Econometric Evaluation Estimator: Evidence from Evaluating a Job Training Programme". In: *The Review of Economic Studies* 64.4, pp. 605–654. doi: 10.2307/2971733.
- Heckman, J. J., L. J. Lochner, and P. E. Todd (2006b). "Chapter 7 Earnings Functions, Rates of Return and Treatment Effects: The Mincer Equation and Beyond". In: *Handbook of the Economics of Education*. Vol. 1. Elsevier, pp. 307–458. doi: 10.1016/S1574-0692(06)01007-5.
- Helbig, M. and R. Nikolai (2015). "Die Unvergleichbaren. Der Wandel der Schulsysteme in den deutschen Bundesländern seit 1949". In: *Book, Verlag Julius Klinkhardt*, pp. 1–383.
- Hendricks, L. and T. Schoellman (2018). "Human Capital and Development Accounting: New Evidence from Wage Gains at Migration". In: *The Quarterly Journal of Economics* 133.2, pp. 665–700. doi: 10.1093/qje/qjx047.
- Hermo, S., M. Päälyssaho, D. Seim, and J. M. Shapiro (2022). "Labor Market Returns and the Evolution of Cognitive Skills: Theory and Evidence*". In: *The Quarterly Journal of Economics* 137.4, pp. 2309–2361. doi: 10.1093/qje/qjac022.
- Hershbein, B. and L. B. Kahn (2018). "Do Recessions Accelerate Routine-Biased Technological Change? Evidence from Vacancy Postings". In: *American Economic Review* 108.7, pp. 1737–1772. doi: 10.1257/aer.20161570.
- Heß, P., S. Janssen, and U. Leber (2023). "The Effect of Automation Technology on Workers' Training Participation". In: *Economics of Education Review* 96, p. 102438. doi: 10.1016/j.econedurev.2023.102438.
- Hoffman, M. and S. V. Burks (2017). *Training Contracts, Employee Turnover, and the Returns from Firm-sponsored General Training*. Working Paper. doi: 10.3386/w23247. eprint: 23247.
- Hoffman, M. and S. Tadelis (2021). "People Management Skills, Employee Attrition, and Manager Rewards: An Empirical Analysis". In: *Journal of Political Economy* 129.1, pp. 243–285. doi: 10.1086/711409.
- Hvidman, U. and H. H. Sievertsen (2021). "High-Stakes Grades and Student Behavior". In: *Journal of Human Resources* 56.3, pp. 821–849. doi: 10.3368/jhr.56.3.0718-9620R2.
- Imbens, G. and Y. Xu (2024). *LaLonde (1986) after Nearly Four Decades: Lessons Learned*. doi: 10.48550/arXiv.2406.00827. eprint: 2406.00827.

- Imbens, G. W. (2004). "Nonparametric Estimation of Average Treatment Effects Under Exogeneity: A Review". In: *The Review of Economics and Statistics* 86.1, p. 26.
- Imbens, G. W. and D. B. Rubin (2015). *Causal Inference for Statistics, Social, and Biomedical Sciences: An Introduction*. Cambridge: Cambridge University Press. doi: 10.1017/CB09781139025751.
- Innocenti, S. and M. Golin (2022). "Human Capital Investment and Perceived Automation Risks: Evidence from 16 Countries". In: *Journal of Economic Behavior & Organization* 195, pp. 27–41. doi: 10.1016/j.jebo.2021.12.027.
- Jacobson, L. S., R. J. LaLonde, and D. G. Sullivan (1993). "Earnings Losses of Displaced Workers". In: *The American Economic Review* 83.4, pp. 685–709. JSTOR: 2117574.
- Jäger, S., C. Roth, N. Roussille, and B. Schoefer (2024). "Worker Beliefs About Outside Options". In: *The Quarterly Journal of Economics*, qjae001. doi: 10.1093/qje/qjae001.
- Jalava, N., J. S. Joensen, and E. Pellas (2015). "Grades and Rank: Impacts of Non-Financial Incentives on Test Performance". In: *Journal of Economic Behavior & Organization* 115, pp. 161–196. doi: 10.1016/j.jebo.2014.12.004.
- James, E., N. Alsalam, J. C. Conaty, and D.-L. To (1989). "College Quality and Future Earnings: Where Should You Send Your Child to College?" In: *The American Economic Review* 79.2, pp. 247–252. JSTOR: 1827765.
- Jedwab, R., P. Romer, A. M. Islam, and R. Samaniego (2023). "Human Capital Accumulation at Work: Estimates for the World and Implications for Development". In: *American Economic Journal: Macroeconomics* 15.3, pp. 191–223. doi: 10.1257/mac.20210002.
- Jensen, R. (2010). "The (Perceived) Returns to Education and the Demand for Schooling". In: *Quarterly Journal of Economics* 125.2, pp. 515–548. doi: 10.1162/qjec.2010.125.2.515.
- Jerrim, J., L. A. Lopez-Agudo, O. D. Marcenaro-Gutierrez, and N. Shure (2017). "What Happens When Econometrics and Psychometrics Collide? An Example Using the PISA Data". In: *Economics of Education Review* 61, pp. 51–58. doi: 10.1016/j.econedurev.2017.09.007.
- Jerrim, J., L. Macmillan, J. Micklewright, M. Sawtell, and M. Wiggins (2018). "Does Teaching Children How to Play Cognitively Demanding Games Improve Their Educational Attainment? Evidence from a Randomized Controlled Trial of Chess

- Instruction in England". In: *Journal of Human Resources* 53.4, pp. 993–1021. doi: 10.3368/jhr.53.4.0516.7952R.
- Jessen, J., L. Kinne, and M. Battisti (2025). "Child Penalties in Labour Market Skills". In: *IZA DP No. 17379*.
- Jones, E. B. and J. D. Jackson (1990). "College Grades and Labor Market Rewards". In: *Journal of Human Resources* 25.2, pp. 253–266.
- Kang, L. S. and S. Sharma (2012). "Managerial Skills and Attitude of MBA and Non-MBA Employees: A Comparison". In: *Management and Labour Studies* 37.2, pp. 77–92. doi: 10.1177/0258042X1203700201.
- Katz, L. F. and K. M. Murphy (1992). "Changes in Relative Wages, 1963-1987: Supply and Demand Factors". In: *The Quarterly Journal of Economics* 107.1, pp. 35–78. doi: 10.2307/2118323.
- Katz, L. F., J. Roth, R. Hendra, and K. Schaberg (2022). "Why Do Sectoral Employment Programs Work? Lessons from WorkAdvance". In: *Journal of Labor Economics* 40.S1, S249–S291. doi: 10.1086/717932.
- Kautz, T., J. J. Heckman, R. Diris, B. ter Weel, and L. Borghans (2014). *Fostering and Measuring Skills: Improving Cognitive and Non-Cognitive Skills to Promote Lifetime Success*. Working Paper 20749. National Bureau of Economic Research. doi: 10.3386/w20749.
- Keith, K. and A. McWilliams (1999). "The Returns to Mobility and Job Search by Gender". In: *Industrial and Labor Relations Review* 52.3, pp. 460–477. doi: 10.2307/2525145. JSTOR: 2525145.
- Kirkeboen, L. J., E. Leuven, and M. Mogstad (2016). "Field of Study, Earnings, and Self-Selection". In: *The Quarterly Journal of Economics* 131.3, pp. 1057–1111. doi: 10.1093/qje/qjw019.
- Kleven, H., C. Landais, and G. Leite-Mariante (2024). "The Child Penalty Atlas". In: *The Review of Economic Studies*, rdae104. doi: 10.1093/restud/rdae104.
- Kleven, H., C. Landais, and J. E. Søgaaard (2019). "Children and Gender Inequality: Evidence from Denmark". In: *American Economic Journal: Applied Economics* 11.4, pp. 181–209. doi: 10.1257/app.20180010.
- KNOKE, . and .ISHIO (1998). "The Gender Gap in Company Job Training". In: *Work and Occupations* 25.2, pp. 141–167. doi: 10.1177/0730888498025002002.
- Koch, A., J. Nafziger, and H. S. Nielsen (2015). "Behavioral Economics of Education". In: *Journal of Economic Behavior & Organization*. Behavioral Economics of Education 115, pp. 3–17. doi: 10.1016/j.jebo.2014.09.005.

- Konings, J. and S. Vanormelingen (2015). "The Impact of Training on Productivity and Wages: Firm-Level Evidence". In: *The Review of Economics and Statistics* 97.2, pp. 485–497. JSTOR: 43556188.
- Kosse, F., T. Deckers, P. Pinger, H. Schildberg-Hörisch, and A. Falk (2020). "The Formation of Prosociality: Causal Evidence on the Role of Social Environment". In: *Journal of Political Economy* 128.2, pp. 434–467. DOI: 10.1086/704386.
- Kostorz, P. (2016). "Bewertungsmaßstäbe und Bezugsnormen bei der Notenvergabe unter der Lupe des Schulrechts – Was ist pädagogisch sinnvoll, was juristisch möglich?" In: *RdJB Recht der Jugend und des Bildungswesens* 64.2, pp. 270–289. DOI: 10.5771/0034-1312-2016-2-270.
- Kraft, M. A. (2020). "Interpreting Effect Sizes of Education Interventions". In: *Educational Researcher* 49.4, pp. 241–253. DOI: 10.3102/0013189X20912798.
- (2023). "The Effect-Size Benchmark That Matters Most: Education Interventions Often Fail". In: *Educational Researcher* 52.3, pp. 183–187. DOI: 10.3102/0013189X231155154.
- Kratz, G. (2021). *Why Use LinkedIn? Here Are 7 Benefits*. Retrieved December 26, 2023. URL: <https://www.flexjobs.com/blog/post/why-use-linkedin-reasons/>.
- Krekel, C., G. Ward, and J.-E. De Neve (2019). *Employee Wellbeing, Productivity, and Firm Performance*. SSRN Scholarly Paper. Rochester, NY. DOI: 10.2139/ssrn.3356581. eprint: 3356581.
- Kristoffersen, J. H. G., M. V. Krægpøth, H. S. Nielsen, and M. Simonsen (2015). "Disruptive School Peers and Student Outcomes". In: *Economics of Education Review* 45, pp. 1–13. DOI: 10.1016/j.econedurev.2015.01.004.
- Kuhn, P. and C. Weinberger (2005). "Leadership Skills and Wages". In: *Journal of Labor Economics* 23.3, pp. 395–436. DOI: 10.1086/430282. JSTOR: 10.1086/430282.
- LaLonde, R. J. (1986). "Evaluating the Econometric Evaluations of Training Programs with Experimental Data". In: *The American Economic Review* 76.4, pp. 604–620. JSTOR: 1806062.
- Länder, F. D. S. Ä. D. B. U. D. (2018). *Mikrozensus 2013, SUF, Version 0*. DOI: 10.21242/12211.2013.00.00.3.1.0.
- Landersø, R. and J. J. Heckman (2017). "The Scandinavian Fantasy: The Sources of Intergenerational Mobility in Denmark and the US". In: *The Scandinavian Journal of Economics* 119.1, pp. 178–230. DOI: 10.1111/sjoe.12219.
- Lazear, E. P. (2001). "Educational Production". In: *The Quarterly Journal of Economics* 116.3, pp. 777–803. JSTOR: 2696418.

- Lechner, M., C. Wunsch, and R. Miquel (2011). "Long-Run Effects of Public Sector Sponsored Training in West Germany". In: *Journal of the European Economic Association* 9.4, pp. 742–784. JSTOR: 25836088.
- Lergetporer, P., K. Wedel, and K. Werner (2023). "Automatability of Occupations, Workers' Labor-Market Expectations, and Willingness to Train". In: *IZA DP No. 16687*.
- Leuven, E. (2005). "The Economics of Private Sector Training: A Survey of the Literature". In: *Journal of Economic Surveys* 19.1, pp. 91–111. doi: 10.1111/j.0950-0804.2005.00240.x.
- Leuven, E. and S. A. Løkken (2020). "Long-Term Impacts of Class Size in Compulsory School". In: *Journal of Human Resources* 55.1, pp. 309–348. doi: 10.3368/jhr.55.2.0217.8574R2.
- Levitt, S. D., J. A. List, S. Neckermann, and S. Sadoff (2016). "The Behavioralist Goes to School: Leveraging Behavioral Economics to Improve Educational Performance". In: *American Economic Journal: Economic Policy* 8.4, pp. 183–219. doi: 10.1257/pol.20130358.
- Lindqvist, E. and R. Vestman (2011). "The Labor Market Returns to Cognitive and Noncognitive Ability: Evidence from the Swedish Enlistment". In: *American Economic Journal: Applied Economics* 3.1, pp. 101–128. JSTOR: 25760248.
- LinkedIn (2024). *About Us*. Retrieved July 15, 2024. URL: <https://news.linkedin.com/about-us#Statistics>.
- (2025). *About Us*. LinkedIn Pressroom. URL: <https://news.linkedin.com/about-us> (visited on 2025).
- Lise, J. and F. Postel-Vinay (2020). "Multidimensional Skills, Sorting, and Human Capital Accumulation". In: *American Economic Review* 110.8, pp. 2328–2376. doi: 10.1257/aer.20162002.
- List, J. A. (2025). *The Experimentalist Looks Within: Toward an Understanding of Within-Subject Experimental Designs*. Working Paper. doi: 10.3386/w33456. eprint: 33456.
- Loprest, P. J. (1992). "Gender Differences in Wage Growth and Job Mobility". In: *The American Economic Review* 82.2, pp. 526–532. JSTOR: 2117456.
- Macdonald, K. (2008). *PV: Stata Module to Perform Estimation with Plausible Values*.
- Maestas, N., K. J. Mullen, D. Powell, T. Von Wachter, and J. B. Wenger (2023). "The Value of Working Conditions in the United States and Implications for the Structure of Wages". In: *American Economic Review* 113.7, pp. 2007–2047. doi: 10.1257/aer.20190846.

- Manning, A. and J. Swaffield (2008). "The Gender Gap in Early-Career Wage Growth". In: *The Economic Journal* 118.530, pp. 983–1024. DOI: 10.1111/j.1468-0297.2008.02158.x.
- Manski, C. F. (2004). "Measuring Expectations". In: *Econometrica* 72.5, pp. 1329–1376. DOI: 10.1111/j.1468-0262.2004.00537.x.
- Mason, H. (2023). *LinkedIn Skills Section—Two Reasons to Update it Today!* Retrieved February 4, 2025. URL: <https://www.jobsearchjourney.com/linkedin-skills/>.
- Maynard, R. A. (1977). "The Effects of the Rural Income Maintenance Experiment on the School Performance of Children". In: *The American Economic Review* 67.1, pp. 370–375. JSTOR: 1815932.
- McCall, B., J. Smith, and C. Wunsch (2016). "Government-Sponsored Vocational Education for Adults". In: *Handbook of the Economics of Education*. Vol. 5. Elsevier, pp. 479–652. DOI: 10.1016/B978-0-444-63459-7.00009-9.
- Merriam-Webster (2025). *Skill*. In *Merriam-Webster.com dictionary*. Retrieved January 28, 2025. URL: <https://www.merriam-webster.com/dictionary/skill>.
- Mincer, J. (1958). "Investment in Human Capital and Personal Income Distribution". In: *Journal of Political Economy* 66.4, pp. 281–302. JSTOR: 1827422.
- Mincer, J. A. (1974). *Schooling, Experience, and Earnings*. NBER.
- Mueller, A. I. and J. Spinnewijn (2023). "Expectations Data, Labor Market, and Job Search". In: *Handbook of Economic Expectations*. Elsevier, pp. 677–713. DOI: 10.1016/B978-0-12-822927-9.00030-6.
- Mueller, A. I., J. Spinnewijn, and G. Topa (2021). "Job Seekers' Perceptions and Employment Prospects: Heterogeneity, Duration Dependence, and Bias". In: *American Economic Review* 111.1, pp. 324–363. DOI: 10.1257/aer.20190808.
- Mummolo, J. and E. Peterson (2019). "Demand Effects in Survey Experiments: An Empirical Assessment". In: *American Political Science Review* 113.2, pp. 517–529. DOI: 10.1017/S0003055418000837.
- Murnane, R. J., J. B. Willett, and F. Levy (1995). "The Growing Importance of Cognitive Skills in Wage Determination". In: *The Review of Economics and Statistics* 77.2, pp. 251–266. DOI: 10.2307/2109863. JSTOR: 2109863.
- Nagler, M., J. Rincke, and E. Winkler (2023). "High-Pressure, High-Paying Jobs?" In: *The Review of Economics and Statistics*, pp. 1–45. DOI: 10.1162/rest_a_01362.
- National Center for Education Statistics (2024). *The Classification of Instructional Programs (CIP)*. Retrieved January 8, 2024. URL: <https://nces.ed.gov/ipeds/cipcode/default.aspx?y=56>.

- Nelissen, J., A. Forrier, and M. Verbruggen (2017). "Employee Development and Voluntary Turnover: Testing the Employability Paradox". In: *Human Resource Management Journal* 27.1, pp. 152–168. doi: 10.1111/1748-8583.12136.
- Niederle, M. and L. Vesterlund (2007). "Do Women Shy Away From Competition? Do Men Compete Too Much?" In: *The Quarterly Journal of Economics* 122.3, pp. 1067–1101. doi: 10.1162/qjec.122.3.1067.
- Nitsche, N. and H. Brckner (2021). "Late, But Not Too Late? Postponement of First Birth Among Highly Educated US Women". In: *European Journal of Population / Revue Européenne de Démographie* 37.2, pp. 371–403. JSTOR: 45381237.
- OECD (2021). *Education at a Glance 2021: OECD Indicators*. Paris: OECD Publishing.
- Olivetti, C., J. Pan, and B. Petrongolo (2024). "The Evolution of Gender in the Labor Market". In: *Handbook of Labor Economics*. Ed. by C. Dustmann and T. Lemieux. Vol. 5. Elsevier, pp. 619–677. doi: 10.1016/bs.heslab.2024.11.010.
- Oreopoulos, P., U. Petronijevic, C. Logel, and G. Beattie (2020). "Improving Non-Academic Student Outcomes Using Online and Text-Message Coaching". In: *Journal of Economic Behavior & Organization* 171, pp. 342–360. doi: 10.1016/j.jebo.2020.01.009.
- Osiander, C. and G. Stephan (2018). "Gerade geringqualifizierte Beschäftigte sehen bei der beruflichen Weiterbildung viele Hürden". In: *LAB Forum*.
- Oster, E. (2019). "Unobservable Selection and Coefficient Stability: Theory and Evidence". In: *Journal of Business & Economic Statistics* 37.2, pp. 187–204. doi: 10.1080/07350015.2016.1227711.
- Oswald, A. J., E. Proto, and D. Sgroi (2015). "Happiness and Productivity". In: *Journal of Labor Economics* 33.4, pp. 789–822. doi: 10.1086/681096.
- Pan, A., B. Seward, and E. Dhuey (2021). *Growing Pains: Parenthood and the Gender Training Gap*. Future Skills Lab Report. University of Toronto.
- Peter, F., C. K. Spiess, and V. Zambre (2021). "Informing Students about College: Increasing Enrollment Using a Behavioral Intervention?" In: *Journal of Economic Behavior & Organization* 190, pp. 524–549. doi: 10.1016/j.jebo.2021.07.032.
- Pfeifer, C. (2016). "Intra-Firm Wage Compression and Coverage of Training Costs: Evidence from Linked Employer-Employee Data". In: *ILR Review* 69.2, pp. 435–454. doi: 10.1177/0019793915610307.
- Pinquart, M., L. P. Juang, and R. K. Silbereisen (2003). "Self-Efficacy and Successful School-to-Work Transition: A Longitudinal Study". In: *Journal of Vocational Behavior* 63.3, pp. 329–346. doi: 10.1016/S0001-8791(02)00031-3.

- Piopiunik, M., G. Schwerdt, L. Simon, and L. Woessmann (2020). "Skills, Signals, and Employability: An Experimental Investigation". In: *European Economic Review* 123, p. 103374. doi: 10.1016/j.euroecorev.2020.103374.
- Pischke, J. (2001). "Continuous Training in Germany". In: *Journal of Population Economics* 14.3, pp. 523–548. JSTOR: 20007779.
- Prenzel, M. et al. (2007). *Programme for International Student Assessment 2003 (PISA 2003)* *Programme for International Student Assessment 2003 (PISA 2003)*. doi: 10.5159/IQB_PISA_2003_V1.
- Prenzel, M. et al. (2010). *Programme for International Student Assessment 2006 (PISA 2006)* *Programme for International Student Assessment 2006 (PISA 2006)*. doi: 10.5159/IQB_PISA_2006_V1.
- Prenzel, M. et al. (2019). *Programme for International Student Assessment 2012 (PISA 2012)* *Programme for International Student Assessment 2012 (PISA 2012)*. doi: 10.5159/IQB_PISA_2012_V5.
- Protsch, P. and H. Solga (2015). "How Employers Use Signals of Cognitive and Noncognitive Skills at Labour Market Entry: Insights from Field Experiments". In: *European Sociological Review* 31.5, pp. 521–532. doi: 10.1093/esr/jcv056.
- Rebollo-Sanz, Y. F. and S. De la Rica (2022). "Gender Gaps in Skills and Labor Market Outcomes: Evidence from the PIAAC". In: *Review of Economics of the Household* 20.2, pp. 333–371. doi: 10.1007/s11150-020-09523-w.
- Resnjanskij, S., J. Ruhose, S. Wiederhold, L. Woessmann, and K. Wedel (2024). "Can Mentoring Alleviate Family Disadvantage in Adolescence? A Field Experiment to Improve Labor Market Prospects". In: *Journal of Political Economy* 132.3, pp. 1013–1062. doi: 10.1086/726905.
- Roodman, D., M. Ø. Nielsen, J. G. MacKinnon, and M. D. Webb (2019). "Fast and Wild: Bootstrap Inference in Stata Using Boottest". In: *The Stata Journal: Promoting communications on statistics and Stata* 19.1, pp. 4–60. doi: 10.1177/1536867X19830877.
- Rosen, S. (1972). "Learning and Experience in the Labor Market". In: *The Journal of Human Resources* 7.3, pp. 326–342. doi: 10.2307/145087. JSTOR: 145087.
- Rosenbaum, P. R. and D. B. Rubin (1983). "The Central Role of the Propensity Score in Observational Studies for Causal Effects". In: *Biometrika* 70.1, pp. 41–55. doi: 10.1093/biomet/70.1.41.
- Roussille, N. (2024). "The Role of the Ask Gap in Gender Pay Inequality". In: *The Quarterly Journal of Economics* 139.3, pp. 1557–1610. doi: 10.1093/qje/qjae004.

- Rubin, D. B. (1987). *Multiple Imputation for Nonresponse in Surveys*. John Wiley & Sons, Ltd. doi: 10.1002/9780470316696.indauth.
- Rubinstein, Y. and Y. Weiss (2006). "Post Schooling Wage Growth: Investment, Search and Learning". In: *Handbook of the Economics of Education*. Ed. by E. Hanushek and F. Welch. Vol. 1. Elsevier, pp. 1–67. doi: 10.1016/S1574-0692(06)01001-4.
- Ruggles, S. et al. (2024). *IPUMS USA: Version 15.0*. [dataset]. Minneapolis, MN.
- Ruhose, J., S. L. Thomsen, and I. Weilage (2019). "The Benefits of Adult Learning: Work-related Training, Social Capital, and Earnings". In: *Economics of Education Review* 72, pp. 166–186. doi: 10.1016/j.econedurev.2019.05.010.
- Rury, D. and S. E. Carrell (2023). "Knowing What It Takes: The Effect of Information about Returns to Studying on Study Effort and Achievement". In: *Economics of Education Review* 94, p. 102400. doi: 10.1016/j.econedurev.2023.102400.
- Ryan, P. (2001). "The School-to-Work Transition: A Cross-National Perspective". In: *Journal of Economic Literature* 39.1, pp. 34–92. JSTOR: 2698454.
- Ryan, R. (2020). *95% of Recruiters Are on LinkedIn Looking for Job Candidates. How to Impress Them*. Retrieved July 15, 2024. Forbes. URL: <https://www.forbes.com/sites/robinryan/2020/09/09/95-of-recruiters-are-on-linkedin-looking-for-job-candidates-how-to-impress-them/>.
- Sachse, K. A. et al. (2012). "IQB-Ländervergleich 2008/2009". In: doi: 10.18452/3126.
- Sanders, C. and C. Taber (2012). "Life-Cycle Wage Growth and Heterogeneous Human Capital". In: *Annual Review of Economics* 4.1, pp. 399–425. doi: 10.1146/annurev-economics-080511-111011.
- Sandgren, S. (2007). "Education and Earnings over the Life Cycle: Longitudinal Age-Earnings Profiles from Sweden". In: doi: 10.7551/mitpress/6051.003.0006.
- Schipolowski, S., N. Haag, F. Milles, S. Pietz, and P. Stanat (2019). *IQB-Bildungstrend 2015*. Humboldt-Universität zu Berlin, Institut zur Qualitätsentwicklung im Bildungswesen. doi: 10.18452/19997.
- Schlosser, A., Z. Neeman, and Y. Attali (2019). "Differential Performance in High Versus Low Stakes Tests: Evidence from the Gre Test". In: *The Economic Journal* 129.623, pp. 2916–2948. doi: 10.1093/ej/uez015.
- Schoner, F., L. Mergele, and L. Zierow (2024). "Grading Student Behavior". In: *Labour Economics* 90, p. 102570. doi: 10.1016/j.labeco.2024.102570.
- Schultz, T. W. (1961). "Investment in Human Capital". In: *The American Economic Review* 51.1, pp. 1–17. JSTOR: 1818907.

- Sloane, C. M., E. G. Hurst, and D. A. Black (2021). "College Majors, Occupations, and the Gender Wage Gap". In: *Journal of Economic Perspectives* 35.4, pp. 223–248. doi: 10.1257/jep.35.4.223.
- Smith, M. (2023). *This Common LinkedIn Mistake Can Hurt Your Chances of Landing a Job Offer—How to Avoid It*. Retrieved January 29, 2025. URL: <https://www.cnbc.com/2023/08/31/career-expert-avoid-this-common-linkedin-mistake-when-job-hunting.html>.
- Speer, J. D. (2020). "Where the Girls Are: Examining and Explaining the Gender Gap in the Nursing Major". In: *Scottish Journal of Political Economy* 67.3, pp. 322–343. doi: 10.1111/sjpe.12234.
- Spence, M. (1973). "Job Market Signaling". In: *The Quarterly Journal of Economics* 87.3, pp. 355–374. doi: 10.2307/1882010. JSTOR: 1882010.
- Spitz-Oener, A. (2006). "Technical Change, Job Tasks, and Rising Educational Demands: Looking Outside the Wage Structure". In: *Journal of Labor Economics* 24.2, pp. 235–270. doi: 10.1086/499972. JSTOR: 10.1086/499972.
- Stantcheva, S. (2023). *How to Run Surveys: A Guide to Creating Your Own Identifying Variation and Revealing the Invisible*. Tech. rep. w30527. National Bureau of Economic Research. doi: 10.3386/w30527.
- Stigler, G. J. (1961). "The Economics of Information". In: *Journal of Political Economy*. doi: 10.1086/258464.
- Stinebrickner, R. and T. R. Stinebrickner (2004). "Time-Use and College Outcomes". In: *Journal of Econometrics*. Higher Education (Annals Issue) 121.1, pp. 243–269. doi: 10.1016/j.jeconom.2003.10.013.
- Streeck, W. (2011). "Skills and Politics: General and Specific". In: *The Political Economy of Collective Skill Formation*. Ed. by M. R. Bussemeyer and C. Trampusch. Oxford University Press, p. 0. doi: 10.1093/acprof:oso/9780199599431.003.0012.
- StudentNews Group (2024). *Degree Abbreviations*. Retrieved January 8, 2024. URL: <https://degree.studentnews.eu/>.
- Sun, L. and S. Abraham (2021). "Estimating Dynamic Treatment Effects in Event Studies with Heterogeneous Treatment Effects". In: *Journal of Econometrics*. Themed Issue: Treatment Effect 1 225.2, pp. 175–199. doi: 10.1016/j.jeconom.2020.09.006.
- Tan, B. J. (2022). "The Consequences of Letter Grades on Labor Market Outcomes and Student Behavior". In: *Journal of Labor Economics* forthcoming. doi: 10.1086/719994.

- Thornton, R. J., J. D. Rodgers, and M. L. Brookshire (1997). "On the Interpretation of Age-Earnings Profiles". In: *Journal of Labor Research* 18.2, pp. 351–365. doi: 10.1007/s12122-997-1043-2.
- Times Higher Education (2019). *Wall Street Journal/Times Higher Education College Rankings 2019*. Retrieved January 8, 2024. URL: <https://www.timeshighereducation.com/rankings/united-states/2019>.
- Tinbergen, J. (1974). "Substitution of Graduate by Other Labour". In: *Kyklos* 27.2, pp. 217–226. doi: 10.1111/j.1467-6435.1974.tb01903.x.
- Topel, R. H. and M. P. Ward (1992). "Job Mobility and the Careers of Young Men". In: *The Quarterly Journal of Economics* 107.2, pp. 439–479. doi: 10.2307/2118478. JSTOR: 2118478.
- Tuch, P. (2000). "Sozialverhalten im Zeugnis – Betragen: Sehr Gut". In: *Deutsches Ärzteblatt*.
- Tyre, P. (2010). "A's for Good Behavior". In: *The New York Times*.
- UCLA Career Center (2024). *Updating Your Profile on LinkedIn: Skills You Should (and Shouldn't) Add*. Retrieved February 5, 2025. URL: <https://career.ucla.edu/blog/2024/10/08/updating-your-profile-on-linkedin-skills-you-should-and-shouldnt-add/>.
- Urabe, M. (2006). "Cultural Barriers in Educational Evaluation: A Comparative Study on School Report Cards in Japan and Germany". In: *International Education Journal* 7.3, pp. 273–283.
- Weidmann, B., J. Vecci, F. Said, D. J. Deming, and S. R. Bhalotra (2024). *How Do You Find a Good Manager?* Working Paper. doi: 10.3386/w32699. eprint: 32699.
- Weinberger, C. J. (2014). "The Increasing Complementarity Between Cognitive and Social Skills". In: *The Review of Economics and Statistics* 96.5, pp. 849–861. JSTOR: 43554962.
- Werner, K. et al. (2022). "Deutsche befürworten Weiterbildung, um mit dem Strukturwandel Schritt zu halten. Ergebnisse des ifo Bildungsbarometers 2022". In: *ifo Schnelldienst* 9.
- Wheeler, L., R. Garlick, E. Johnson, P. Shaw, and M. Gargano (2022). "LinkedIn(to) Job Opportunities: Experimental Evidence from Job Readiness Training". In: *American Economic Journal: Applied Economics* 14.2, pp. 101–125. doi: 10.1257/app.20200025.
- Wiswall, M. and B. Zafar (2015). "Determinants of College Major Choice: Identification Using an Information Experiment". In: *The Review of Economic Studies* 82.2, pp. 791–824. doi: 10.1093/restud/rdu044.

- Wiswall, M. and B. Zafar (2021). "Human Capital Investments and Expectations about Career and Family". In: *Journal of Political Economy* 129.5, pp. 1361–1424. doi: 10.1086/713100.
- Woessmann, L. (2016). "The Importance of School Systems: Evidence from International Differences in Student Achievement". In: *Journal of Economic Perspectives* 30.3, pp. 3–32. doi: 10.1257/jep.30.3.3.
- (2025). "Skills and Earnings: A Multidimensional Perspective on Human Capital". In: *Annual Review of Economics* 17.
- Wright, M. N. and A. Ziegler (2017). "Ranger: A Fast Implementation of Random Forests for High Dimensional Data in C++ and R". In: *Journal of Statistical Software* 77, pp. 1–17. doi: 10.18637/jss.v077.i01.
- Yeager, D. S. and C. S. Dweck (2020). "What Can Be Learned from Growth Mindset Controversies?" In: *American Psychologist* 75.9, pp. 1269–1284. doi: 10.1037/amp0000794.
- YourDictionary (2022). *Degree Abbreviations*. Retrieved January 8, 2024. URL: <https://www.yourdictionary.com/articles/degree-abbreviations>.
- Zafar, B. (2013). "College Major Choice and the Gender Gap". In: *Journal of Human Resources* 48.3, pp. 545–595. doi: 10.3368/jhr.48.3.545.
- Zimmermann, K. F. et al. (2013). "Youth Unemployment and Vocational Training". In: *Foundations and Trends® in Microeconomics* 9.1-2, pp. 1–157. doi: 10.1561/07000000058.