Differences in returns to vocational and general education: evidence from a differentiated schooling system *

Guillem Foucault† François Poinas‡

This draft: October 27, 2026

Abstract

The literature on returns to education has focused on years of schooling and fields of study, and less on whether post-secondary education is vocational or academic. We study how this choice affects occupations and wages in Germany. We estimate a dynamic semi-structural model of completed post-secondary education and first occupation choice, using the German Socio-Economic Panel. Occupations are described by their cognitive, interpersonal and manual task content, and selection runs through latent cognitive and non-cognitive abilities. Cognitive ability strongly predicts post-secondary attainment, especially university, while non-cognitive ability plays no role once cognitive ability is accounted for. Conditional on abilities and occupational tasks, an apprenticeship raises first-job hourly wages by 11% relative to basic education, and a university degree by 45%. Vocational degrees move workers towards manual and cognitive occupations and away from interpersonal ones. This occupational sorting explains only a modest part of the wage returns to education (8 to 20%), and almost none of the gap between university and apprenticeship graduates, which arises within occupations.

Keywords: Factor model, Vocational Education, Task-based approach
JEL Codes: I26, I28, J24

*We thank Olivier De Groote for his valuable comments. We acknowledge funding from the French National Research Agency (ANR) under the grant RECHC - ANR-20-CE26-0011-01.

†Toulouse School of Economics. guillem.foucault@tse-fr.eu. gfill.github.io.

‡Toulouse School of Economics. francois.poinas@tse-fr.eu. sites.google.com/site/francoispoinas.

1 Introduction

Schooling attainments differ not only by the quantity of schooling (accumulated number of years of education), but also by the type of schooling. In post-secondary education, one aspect of the qualitative dimension, that is well studied in the literature, concerns the fields of study (science, humanities, law, etc.). The empirical evidence in favour of heterogeneity in returns across fields of study is relatively important (Altonji et al., 2016), suggesting the presence of substantial differences in human capital acquisition across fields.

Another important qualitative dimension of post-secondary education is the nature of education programs (vocational or general). This dimension has attracted less attention, even though this distinction is an important feature of many educational systems, especially in continental Europe. This aspect of post-secondary education is an important element to be considered, from a policy perspective, to identify educational programs that maximize the skills needed in the labour market and understand the sources of mismatch between training and occupations.

This paper studies how the nature of post-secondary education (academic vs. vocational) shapes occupational choices and wages in Germany, and through which channels. We ask three related questions. First, how do individuals sort into academic and vocational degrees on their cognitive and non-cognitive abilities? Second, how do these degrees change the task content of the occupations graduates enter? Third, how much of the wage return to each degree arises within occupations, and how much through access to occupations whose task content is better paid? Answering these questions helps understand the role that post-secondary programmes play in providing the skills rewarded in the labour market. In particular, it tells us whether programmes with training specific to a small range of occupations pay off mainly by channelling graduates into particular occupations, or by raising their

wages within a given occupation. The German context is particularly relevant to address this question, with a well-developed vocational-oriented system that combines classroom lectures with on-the-job training. The prevalence of these programs is usually put forward as an explanation for the low youth unemployment rate in Germany[1].

The literature on returns to education have focused on dealing with unobserved ability (Heckman and Rubinstein, 2001). But human capital is made of multidimensional skills: individuals are endowed and acquire skills of different nature (cognitive, manual, interpersonal). Not accounting for this multidimensionality can lead to confusion in interpreting returns to schooling (Heckman, Humphries, et al., 2018). Occupations are heterogeneous in the skill intensity needed to perform the job: individuals with a high cognitive ability might struggle to complete manual tasks, and vice-versa. Therefore, choosing wisely the post-secondary education is important to develop individuals' abilities and find an occupation that matches the job skill requirements.

Our empirical approach consists in estimating a dynamic semi-structural model of completed post-secondary education and first occupation choice. Students enter the model sorted into a lower secondary track (*Hauptschule*, *Realschule* or *Gymnasium*). We treat the track as an initial condition that shifts preferences and restricts the set of attainable degrees. Students then choose their highest completed level of post-secondary education among five levels: basic education, the upper-secondary *Abitur*, apprenticeship, higher vocational education and university. Once they leave education, they choose an occupation, which determines their wage. A key element of the model is that occupations are characterised by their task content, summarised in three dimensions (cognitive, interpersonal and manual) built from the BIBB/BAuA task survey following Yamaguchi (2012). Individuals value these tasks

[1]In 2019, 4.7% of youngsters were unemployed compared to the 12.1% EU27 average using the 15-29 years old definition, 6.2% vs 15.6% for the 15-24 years old definition (Eurostat, 2025).

differently depending on their education, background and abilities, and the wage equation prices the task content of the chosen occupation.

Selection into education and occupations runs through two latent factors, cognitive and non-cognitive ability. We identify them from a measurement system of nine indicators observed at age 17, before post-secondary choices are made: three IST 2000R cognitive test scores, three school grades and three non-cognitive indices (conscientiousness, locus of control and self-esteem). The measurement system conditions on the lower secondary track, so the factors capture ability net of the track. We estimate the model in two steps. The first step recovers the factor distribution from the measurement system by full-information maximum likelihood, which lets us keep individuals with missing measurements. The second step estimates the education, occupation and wage equations jointly, integrating over each individual's posterior distribution of abilities, and corrects standard errors for first-stage estimation error following Murphy and Topel (1985).

The model allows us to quantify the wage returns to vocational and academic degrees net of ability sorting. It also allows us to separate the part of these returns that arises within occupations from the part that arises through sorting into occupations with different task content. Symmetrically, it lets us decompose wage differences across occupations into what an occupation pays a given worker for its task content and the composition of its workforce in terms of education, abilities and background.

We find strong sorting on cognitive ability into post-secondary education, especially into academic degrees. A one standard deviation increase in cognitive ability raises the log-odds of completing university rather than basic education by 1.6, more than twice its effect on apprenticeship or higher vocational education. Non-cognitive ability plays no role in the education choice once cognitive ability is accounted for. The lower secondary track matters

even more than ability. Compared with *Realschule* students, *Gymnasium*-track students have log-odds of completing the *Abitur* or university, rather than an apprenticeship, that are higher by about 4.9 and 3.7. These are by far the largest effects on the education choice. Occupational sorting is driven mainly by gender and education. Vocational degrees move individuals towards occupations rich in manual and cognitive tasks and away from interpersonal ones. Latent abilities add little once education is controlled for: ability sorting into occupations operates mostly through the education choice.

Wage returns to post-secondary education rise with the level of the degree. Conditional on the task content of the occupation and on abilities, an apprenticeship raises first-job hourly wages by 11% relative to basic education, a higher vocational degree by 28% and a university degree by 45%. The lower secondary track has a wage effect of its own, whatever the final degree: having attended *Gymnasium* raises wages by 17% relative to *Hauptschule*. The track therefore shapes outcomes both through the degrees it leads to and directly. Occupations are paid for their task content: moving to an occupation whose cognitive content is one standard deviation higher, measured across first jobs, raises wages by about 10%, while interpersonal content is penalised. Conditional on education and occupation, the latent abilities have small and insignificant direct effects on wages, so cognitive ability affects wages mainly through the education choice. Decomposing the wage returns, we find that sorting into occupations with better-paid task content accounts for 8% of the university premium, 12% of the higher vocational premium and 20% of the apprenticeship premium, and for almost none of the gap between university and apprenticeship graduates. Returns to academic and vocational degrees therefore arise mostly within occupations, not from access to different occupations. Across occupations, what an occupation pays a given worker for its task content accounts for slightly less than 30% of the variance in mean first-job wages. The composition of its workforce accounts for about 40%, mostly through education and background.

Our first contribution is to the task-based literature, which characterises occupations by the tasks they involve rather than by discrete categories. Gathmann and Schönberg (2010) were the first to use task information to build a continuous measure of the skill requirements of occupations. They estimate returns to skills and find that human capital is more portable than previously considered. Yamaguchi (2012), using a similar approach, estimates skill transferability and finds that cognitive-intensive occupations are more transferable than manual-intensive ones. This literature takes workers' skills as given and focuses on their returns and their transferability across occupations. We move one step upstream: we model how post-secondary education choices, driven by latent abilities and the lower secondary track, determine the task content of the occupations individuals enter and the wage they earn for it. We show that education, rather than latent abilities, drives sorting across occupations' task content: abilities shape occupational outcomes mostly through the education choice.

The link between post-secondary education and occupation has mostly been studied along the field-of-study dimension: the returns to working in an occupation related to one's major (Kinsler and Pavan, 2015; Lemieux, 2014) or, in the German context, to one's training occupation (Eckardt, 2026). Less is known about how the academic–vocational divide in post-secondary education shapes occupational sorting, net of selection on abilities. We contribute by studying this link, characterising occupations by their task content. We show that, although vocational and academic graduates sort into occupations with different task content, this sorting explains little of the wage gap between them.

Our work also contributes to the extensive literature estimating returns to education accounting for unobserved ability. First attempts included finding proxies for cognitive abilities (Castex and Dechter, 2014). More recent literature focuses instead on the dimension of unobserved ability, acknowledging its multidimensionality (Heckman and Vytlacil, 2001;

Heckman, Humphries, et al., 2018; Ashworth et al., 2021).

More concretely, we measure the impact of the qualitative nature of education on wages. A vast literature focuses on returns to field of study (Arcidiacono, 2004; Beffy et al., 2012), but the nature of education is relatively unexplored with a few exceptions (Belzil and Poinas, 2018; De Groote and Declercq, 2021). We explore this relationship in the German context, where vocational training is an important part of the education system.

This paper proceeds in the following way. Section 2 introduces the institutional background. Section 3 describes the data and presents descriptive statistics. Section 4 defines the model of education and occupational choices. Section 5 discusses identification and Section 6 the estimation procedure. Section 7 presents the results, including the decomposition of wage returns into education and occupation channels. Finally, Section 8 concludes and presents next steps.

2 Institutional Setting

The German educational system differs from most other European countries (except for Austria and Switzerland, with a similar organisation) in two main dimensions. The first particularity is that students are split into different types of secondary school according to their academic ability at the end of primary school, typically after grade 4, based on a teacher recommendation. Three main tracks coexist. *Hauptschule* (grades 5 to 9) provides basic general education and prepares students for vocational training. *Realschule* (grades 5 to 10) provides a more extensive general education. Its students have similar prospects as their *Hauptschule* counterparts, but those with good grades can switch into an academic track afterwards (Dustmann et al., 2017; Biewen and Tapalaga, 2017). Finally, *Gymnasium* provides intensified general education and ends after grade 12 or 13 with the *Abitur*, the

qualification required to enrol at university. The length of *Gymnasium* varies across Federal States and cohorts: most states shortened it from nine to eight years during the 2000s, and the cohorts we study straddle this reform. Some students attend comprehensive schools that offer the three tracks within a single school; we assign them to a track according to the certificate they obtain or aim for.

Secondly, it puts a high emphasis on vocational training. Upon completing compulsory schooling, students can apply for an apprenticeship in the "dual" system, which combines on-the-job training in a firm with part-time classes at a vocational school. Access is not formally conditioned on a school certificate: firms select their apprentices, and a growing share of them hold an *Abitur* (Bundesinstitut für Berufsbildung, 2019). Apprenticeship is the most common option: 65% of non-academic students opted for it in 2018 (European Commission, 2018). It usually lasts three years, during which apprentices receive a training allowance from the employer. The remaining 35% of students attend full-time vocational schools, which combine classroom teaching with work placements and are common in health and social care occupations (Innengruppe Bildungsberichterstattung, 2024).

Beyond vocational training, two higher levels of education are available. Higher vocational education comprises universities of applied sciences (*Fachhochschulen*), which are open to students with an *Abitur* or a more applied entrance qualification, and advanced vocational degrees such as master craftsman, which build on an apprenticeship and some years of work experience. Academic universities require the *Abitur*. Since we model the highest degree completed rather than the sequence of transitions leading to it (Section 4), we summarise these options into the five education levels of our analysis: basic education, *Gymnasium*, apprenticeship, higher vocational education and university.

3 Data

We use the German Socioeconomic Panel (GSOEP), a longitudinal survey of approximately 15,000 households since 1984 administered by the German Institute for Economic Research (DIW Berlin). We use the individual, youth and cognitive questionnaires. One of the most important features from our perspective is that, between 2006 and 2025, 3,931 individuals undertook a standardised cognitive IST 2000R test developed by Amthauer et al. (2001). This test is taken around the age of 17 (up to 19 in the first, 2006 wave) and seeks to summarise general cognitive components including verbal reasoning, numerical abilities, and figural abilities.

We merge questions related to locus of control, self-esteem and conscientiousness from the youth questionnaire, answered at age 17 like the cognitive test. They are comprised of a series of questions measured on the Likert scale (see Table 1). We also make use of self-reported high school grades in German, Maths and a Foreign Language. Furthermore, it contains biometric measures such as height and weight, and, only for a more restricted sample, grip strength measures, which are strongly correlated with general health conditions. More generally, it includes socio-economic information about respondents, and detailed education and labour market outcomes. For instance, 89% of university graduates were in *Gymnasium* at the age of 16, and 56% of apprentices come from the *Realschule.*

Table 1: Descriptive statistics, estimation sample

	Mean	SD	Min	Max
Demographics				
Female	0.47	–	–	–
Rural	0.38	–	–	–
Monthly household income (1,000 EUR)	3.57	2.18	0.00	20.00
Lower secondary track				
Hauptschule	0.16	–	–	–
Realschule	0.35	–	–	–
Gymnasium	0.48	–	–	–
Completed education				
Basic education	0.06	–	–	–
Gymnasium (upper secondary)	0.06	–	–	–
Apprenticeship	0.41	–	–	–
Higher vocational	0.21	–	–	–
University	0.27	–	–	–
Cognitive tests (IST 2000R)				
Word analogies (0–20)	8.22	3.60	0.00	20.00
Arithmetic (0–20)	12.84	4.80	0.00	20.00
Matrices (0–20)	9.93	3.35	0.00	19.00
School grades (1–6, higher = better)				
German ($n = 979$)	4.04	0.83	1.00	6.00
First foreign language ($n = 957$)	3.99	0.93	2.00	6.00
Mathematics ($n = 979$)	3.99	1.03	2.00	6.00
Personality (1–7)				
Conscientiousness ($n = 903$)	4.92	1.11	1.00	7.00
Locus of control ($n = 905$)	4.95	0.84	2.00	7.00
Self-esteem ($n = 553$)	5.34	1.35	1.00	7.00
Labour market (first job)				
Real hourly gross wage (2020 EUR) ($n = 1{,}114$)	12.04	5.24	1.68	29.38
Log real hourly gross wage ($n = 1{,}114$)	2.38	0.50	0.52	3.38
Observations	1,137			

Notes: One observation per individual. For binary variables only the share is reported. Variables with missing values are summarised over observed values (n in parentheses). The factor model integrates missing measurements out (FIML); individuals without a first-job wage enter the education and occupation choices only.

48% of students follow the academic track, followed by *Realschule* (35%) and *Hauptschule* (16%). These initial tracks partly determine the type of post-secondary education students

can access (see Figure 1 below).

Figure 1: Education transitions: lower secondary track to completed education

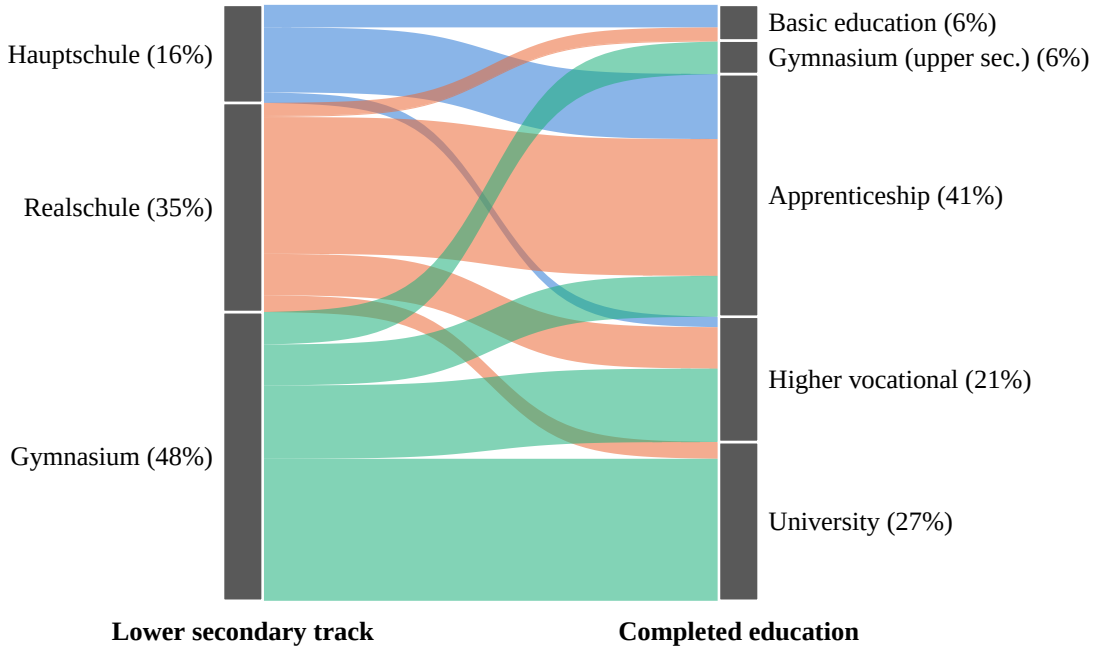


Figure 1 shows that some transitions are almost never observed. *Hauptschule* students do not obtain the *Abitur*, and therefore cannot enter university. *Gymnasium* students almost always complete at least the upper secondary degree, so they do not leave with basic education only. We treat these transitions as infeasible: each lower secondary track defines its own set of attainable levels in the education choice model (Section 4.1). *Hauptschule* students choose between basic education, apprenticeship and higher vocational education. *Realschule* students can reach all five levels, and *Gymnasium* students all levels but basic education. Eight individuals in the estimation sample report an excluded transition, all from *Hauptschule* to university, and we drop them. These restrictions also determine the reference categories in the education choice estimates (see Section 7.2).

The data regarding tasks performed in occupations is gathered from the German Qualification and Career Survey, which is conducted jointly by the Federal Institute for Vocational Education and Training (BIBB) and the German Federal Institute for Occupational Safety and Health (BAuA). We use the latest available survey of year 2018. Surveyed workers are asked about 18 different tasks performed at the workplace, in which they have to answer whether they perform the task "frequently", "sometimes" or "never". This survey has been widely used to measure the task content of German occupations (Spitz-Oener, 2006; Gathmann and Schönberg, 2010). Gathmann and Schönberg (2010), and Eckardt (2026) after them, keep the full vector of activities and measure the distance between occupations by their angular separation, without grouping activities into categories. Since our model requires interpretable task intensities that interact with individual abilities, we instead summarise each occupation in a three-dimensional space: "Cognitive", "Manual" and "Interpersonal". Each activity is assigned to exactly one category, following the task types of Dengler et al. (2014) collapsed to three dimensions (see Appendix Section A.2). Within each category, we employ a "directed" Principal Components Analysis (PCA) approach as in Yamaguchi (2012): we retain the first principal component of the occupation-level activity frequencies. As a robustness check, we use the task shares of Spitz-Oener (2006), i.e. the fraction of an individual's activities that fall into each category, averaged at the occupation level.

Labour market outcomes considered are occupational decision and hourly wage. Given the cognitive test was introduced in 2006, individuals are at most 38 years old in 2025, the last available wave, and most are much younger. Therefore, we consider the first occupation and associated wage. The first job is the first year after leaving education in which the individual is in full- or part-time employment (excluding apprenticeships and marginal jobs), reports positive earnings, works at least 10 contracted hours per week and reports an occupation. Occupation and wage are both taken from that year. The hourly wage is gross monthly

labour earnings divided by contracted hours, deflated to 2020 euros with the consumer price index. We trim the 1st and 99th percentiles of the hourly wage distribution.

Our sample consists of individuals that take the cognitive test (IST 2000R) and whom we observe in a first job. Later test cohorts therefore enter the sample only if they have left education and started working by 2025. We do not require the other measurements to be observed. School grades are missing for 14–16% of the estimation sample, conscientiousness and locus of control for 20–21%, and self-esteem for 51%; only 444 individuals report all nine measurements. Rather than dropping these individuals, we integrate the missing measurements out of the first-stage likelihood (full-information maximum likelihood, see Section 6.1). After dropping the eight excluded transitions, the estimation sample comprises 1,137 observations: 187 *Hauptschule*, 399 *Realschule* and 551 *Gymnasium* students. Of these, 1,114 report an hourly wage in their first job; the other 23 enter the education and occupation choices only.

4 Model

We develop a dynamic semi-structural model of schooling decisions and first labour market outcomes. The goal is to estimate wage returns amongst vocational and academic students while controlling for potential multidimensional ability bias.

Students enter the model having been sorted into a lower secondary track, which we treat as an initial condition. They then choose their highest completed level of post-secondary education. Once they leave the educational system, individuals pick an occupation $o \in \mathcal{O}$ so as to maximise their utility, which determines their associated wage w.

4.1 Post-Secondary Education

We model the highest level of education an individual completes after compulsory schooling, $D_i \in \mathcal{J}$, rather than the sequence of enrolment decisions that leads to it. The set $\mathcal{J}$ has five levels: *basic education* (no qualification beyond the lower secondary degree), *Gymnasium* (the upper secondary *Abitur* without any further degree), *apprenticeship* (dual or school-based vocational training), *higher vocational* education (*Fachhochschule* and other higher vocational degrees), and *university*. Working with completed education keeps the model tractable given the size of our sample, and it is the level of education that is relevant for the labour market outcomes we study.

Individuals choose the level j that maximises their utility. The utility of each level depends on the lower secondary track attended at age 16, $\mathbf{S_i}$ (dummies for *Hauptschule*, *Realschule* and *Gymnasium*), on observed characteristics $\mathbf{X_i}$ (gender, household income and a rural dummy), and on cognitive and non-cognitive abilities θ_i^c and θ_i^{nc}, which are unobserved by the econometrician. We assume these components to be additively separable:

$$U_i^j = \beta_0^j + \mathbf{S_i}'\beta_{\mathbf{1}}^{\mathbf{j}} + \mathbf{X_i}'\beta_{\mathbf{2}}^{\mathbf{j}} + \alpha^{\mathbf{c,j}}\theta_{\mathbf{i}}^{\mathbf{c}} + \alpha^{\mathbf{nc,j}}\theta_{\mathbf{i}}^{\mathbf{nc}} + \mathbf{e_i^j}, \qquad \mathbf{j} \in \mathcal{J}(\mathbf{S_i}), \tag{1}$$

where β_0^j is a level-specific constant, $\alpha^{c,j}$ and $\alpha^{nc,j}$ are the factor loadings of the two abilities, and e_i^j is a Gumbel distributed utility shock. Basic education is the reference alternative.

The lower secondary track is an initial condition. Students are sorted into tracks around age ten, before any of the choices we model and before abilities are measured at age 17. It therefore enters Equation 1 in two ways. First, it shifts the utility of each level through $\boldsymbol{\beta}_1^j$. Conditioning on it also helps to interpret the latent factors as abilities rather than as the

product of the track itself. Second, it restricts the set of attainable levels $\mathcal{J}(\mathbf{S}_i) \subseteq \mathcal{J}$, as observed in Figure 1:

$$\begin{aligned}
\mathcal{J}(\textit{Hauptschule}) &= \{\text{basic, apprenticeship, higher vocational}\},\\
\mathcal{J}(\textit{Realschule}) &= \mathcal{J},\\
\mathcal{J}(\textit{Gymnasium}) &= \{\textit{Gymnasium}, \text{apprenticeship, higher vocational, university}\}.
\end{aligned}$$

Hauptschule students do not obtain the *Abitur* or a university degree. *Gymnasium* students complete at least the upper secondary degree, so basic education is not in their choice set. These restrictions have two consequences for identification, since the multinomial logit only identifies utility differences within a choice set. Among the alternatives unavailable to *Hauptschule* students (*Gymnasium* and university), the effect of the lower secondary track is identified relative to *Realschule* only. Furthermore, because the choice set of *Gymnasium* students excludes the reference alternative, we normalise the *Gymnasium*-track coefficient to zero for apprenticeship, which is available to every track.

4.2 Task-Based Occupation Choice

Throughout the model we distinguish two types of educational variables. Lower secondary schooling $\mathbf{S}_i$ — the track attended in compulsory schooling (Hauptschule, Realschule, or Gymnasium) — is an *initial condition* determined prior to any of the choices modelled here. Completed post-secondary education $\mathbf{P}_i$ — dummies for the level $D_i \in \mathcal{J}$ chosen in Section 4.1, with basic education as the reference — is endogenous and enters as a control in subsequent equations.

Occupational utility is derived from the task content of an occupation interacted with individual ability. Each occupation $o \in \mathcal{O}$ is characterised by a vector $\mathbf{T}_o = (T_o^c, T_o^{ip}, T_o^m)$,

corresponding to its cognitive, interpersonal and manual task intensities, obtained via a Principal Component Analysis of task survey data (see Appendix Section A.2).

The utility from choosing occupation o is given by:

$$\begin{aligned} U_{io} = {} & T_o^c(\delta^c + \Gamma_{c,c}\,\theta_i^c + \Gamma_{c,nc}\,\theta_i^{nc} + \mathbf{P}_i'\boldsymbol{\phi}^c + \mathbf{X}_i'\boldsymbol{\psi}^c) \\ & + T_o^{ip}\big(\delta^{ip} + \Gamma_{ip,c}\,\theta_i^c + \Gamma_{ip,nc}\,\theta_i^{nc} + \mathbf{P}_i'\boldsymbol{\phi}^{ip} + \mathbf{X}_i'\boldsymbol{\psi}^{ip}\big) \\ & + T_o^m(\delta^m + \Gamma_{m,c}\,\theta_i^c + \Gamma_{m,nc}\,\theta_i^{nc} + \mathbf{P}_i'\boldsymbol{\phi}^m + \mathbf{X}_i'\boldsymbol{\psi}^m) + \varepsilon_{io}, \end{aligned} \tag{2}$$

where δ^k is the base effect of task dimension k on utility, common to all individuals; $\boldsymbol{\Gamma} = (\Gamma_{k,l})$ is a 3×2 matrix of factor loadings, where $\Gamma_{k,l}$ captures how ability $l \in \{c, nc\}$ moderates preferences for task k; $\boldsymbol{\phi}^k$ and $\boldsymbol{\psi}^k$ are vectors of coefficients on post-secondary education $\mathbf{P}_i$ and background characteristics $\mathbf{X}_i$ interacted with task k; θ_i^c and θ_i^{nc} are latent cognitive and non-cognitive abilities; and ε_{io} is an i.i.d. Gumbel-distributed shock. We leave $\boldsymbol{\Gamma}$ unrestricted, so that both abilities may affect the valuation of every task. In the occupation and wage equations, $\mathbf{X}_i$ also includes the lower secondary track dummies $\mathbf{S}_i$.

Since this is a conditional logit model, individual-specific variables that do not vary across alternatives are not identified separately from alternative-specific constants. Accordingly, $\mathbf{P}_i$, $\mathbf{X}_i$, and $\boldsymbol{\theta}_i$ enter utility *exclusively* through their interaction with the occupation-varying task intensities $\mathbf{T}_o$. The base coefficients $\boldsymbol{\delta}$ capture average preferences for each task type, while $\boldsymbol{\Gamma}$, $\boldsymbol{\phi}^k$, and $\boldsymbol{\psi}^k$ recover how abilities, education, and background translate into differential sorting across task-intensive occupations.

Individual i then chooses $o_i = \arg\max_{o \in \mathcal{O}} U_{io}$.

4.3 Wages

Wages depend not only on individual characteristics and education, but also on the task content of the chosen occupation. We specify the log wage equation as:

$$\log(w_i) = \pi_0 + \mathbf{P}_i'\boldsymbol{\pi}_P + \mathbf{X}_i'\boldsymbol{\pi}_X + \rho^c T^c_{o_i} + \rho^{ip} T^{ip}_{o_i} + \rho^m T^m_{o_i} + \alpha^c_w \theta^c_i + \alpha^{nc}_w \theta^{nc}_i + u_i, \tag{3}$$

where:

- $\mathbf{X}_i$ includes the lower secondary track dummies $\mathbf{S}_i$, as in Equation 2,
- $T^c_{o_i}$, $T^{ip}_{o_i}$, $T^m_{o_i}$ are the task intensities of the occupation o_i chosen by individual i,
- ρ^c, ρ^{ip}, ρ^m are task-specific returns in the wage equation,
- $u_i \sim \mathcal{N}(0, \sigma^2_w)$ is an i.i.d. wage shock.

This structure introduces wage heterogeneity both across and within occupations via their task profiles. It allows us to test whether task content (conditional on education and ability) is rewarded in the labour market. Wages are not separately indexed by occupation, but derive their heterogeneity through the observed task bundle and latent productivity.

4.4 Factors

If abilities were directly observable, we could have directly included them in our model alongside the rest of covariates. A branch of the structural literature treats the factors as a nuisance variable without interpreting them employing a "random effects" approach. Instead, as in Heckman, Stixrud, et al. (2006), we postulate the existence of two underlying

factors representing latent cognitive and non-cognitive ability.

We define a measurement system of latent abilities with $\mathbf{M}$ denoting a vector of measurements. Measurements are assumed to be separable in observables, latent abilities and a normal idiosyncratic shock. Denoting M_m the mth measurement,

$$M_{i,m} = \beta_{0,m} + \mathbf{S_i'}\beta_{\mathbf{1,m}} + \mathbf{X_i}'\beta_{\mathbf{2,m}} + \alpha^{\mathbf{c}}_{\mathbf{m}}\theta^{\mathbf{c}}_{\mathbf{i}} + \alpha^{\mathbf{nc}}_{\mathbf{m}}\theta^{\mathbf{nc}}_{\mathbf{i}} + \mathbf{e_{i,m}}, \tag{4}$$

where $\beta_{0,m}$ is the intercept of measurement m, $\mathbf{S_i}$ controls for the type of mandatory high school track, $\mathbf{X_i}$ is a covariate vector including gender, household income and a rural dummy; θ_i^c and θ_i^{nc} denote the cognitive and non-cognitive latent factors, respectively, and $e_{i,m} \sim \mathcal{N}(0, \sigma^2_{M_m})$ is an idiosyncratic shock, assumed mutually independent across measurements and independent of observables and latent abilities. We assume

$$\mathbb{E}[e_{i,m}] = 0;\ e_{i,m} \perp\!\!\!\perp e_{j,m} \text{ for } i \neq j$$

Controlling for mandatory high school is crucial given the nature of German differentiated schooling system. Since grade five, kids are sorted into different academic paths according to their ability and ambition (as explained in Section 2). Therefore, it is natural to exhibit ability sorting, with *Gymnasium* concentrating the individuals with the highest cognitive abilities on average. Parental background can also influence both cognitive and non-cognitive abilities.

Stacking the nine measurement equations,

$$\mathbf{M} = \begin{pmatrix} M_1 \\ \vdots \\ M_9 \end{pmatrix} = \begin{pmatrix} g_1(\mathbf{S}, \mathbf{X}, \boldsymbol{\theta}, \mathbf{e_1}) \\ \vdots \\ g_9(\mathbf{S}, \mathbf{X}, \boldsymbol{\theta}, \mathbf{e_9}) \end{pmatrix}.$$

Test scores and school grades load only on the cognitive factor, while the non-cognitive measurements load on both factors. These restrictions, together with the normalisations required for identification, are discussed in Section 5.1 and summarised in Table 2.

5 Identification

This section discusses which restrictions identify the model. Identification proceeds in two steps, mirroring the estimation in Section 6. First, the measurement system identifies the distribution of the latent factors and pins down their location, scale and rotation. Second, given the factor distribution, the joint distribution of education, occupation and wages identifies the parameters of the outcome equations, including their factor loadings.

5.1 Measurement System

The identification of the measurement system follows standard arguments (Anderson and Rubin, 1956; Carneiro et al., 2003; Heckman, Stixrud, et al., 2006; Cunha et al., 2010). We only state the assumptions and normalisations we impose, and leave the formal argument to Appendix B. Let $\boldsymbol{\theta}_i = (\theta_i^c, \theta_i^{nc})'$. We assume that

(A1) $\boldsymbol{\theta}_i \perp\!\!\!\perp (\mathbf{S}_i, \mathbf{X}_i)$: the factors are abilities net of the lower secondary track and of family background;

(A2) the measurement errors e_i^m are mutually independent and independent of $(\boldsymbol{\theta}_i, \mathbf{S}_i, \mathbf{X}_i)$;

(A3) $\theta_i^c \perp\!\!\!\perp \theta_i^{nc}$.

Under these assumptions, the covariances between the measurements, net of observables, depend only on the factor loadings and the factor variances. Factors have no natural location, scale or orientation, which we fix as follows.

Location and scale. We normalise $\mathbb{E}[\boldsymbol{\theta}_i] = \mathbf{0}$; in the outcome equations, the mean of the factors is absorbed by the constants. We fix the scale of each factor by normalising one of its loadings to one: the cognitive factor is measured in units of the word analogy test, the non-cognitive factor in units of conscientiousness. This also fixes the sign of each factor. The factor variances are free parameters. The choice of the anchoring measure only affects the unit of the factors. The tables report loadings per unit of the factor, together with the factor standard deviations, and in the text we express all effects of ability per standard deviation of the factor, which does not depend on the anchoring measure.

Rotation. With two factors, two additional restrictions are needed to pin down their rotation (Anderson and Rubin, 1956). First, the six cognitive measurements (three IST test scores and three school grades) are *dedicated* to the cognitive factor. Second, the factors are independent, (A3). The non-cognitive measurements load on both factors: they are self-reported, and the answers of more able students may differ for reasons unrelated to personality. Table 2 summarises the resulting loading structure.

Table 2: Factor loading restrictions

Measures	θ^c	θ^{nc}
IST 2000R cognitive test		
Word Analogy	1	0
Arithmetic	x	0
Spatial Reasoning (matrices)	x	0
Youth survey: school grades		
German	x	0
Foreign Language	x	0
Mathematics	x	0
Youth survey: non-cognitive		
Conscientiousness	x	1
Locus of Control	x	x
Self-esteem	x	x

Notes: x denotes a freely estimated loading, 1 a loading normalised to one and 0 a loading restricted to zero. Factors are independent, with mean zero and free variances.

With correlated factors, a non-cognitive measurement that covaries with the cognitive ones could reflect either a cross-loading or a correlation between the factors, and the data could not tell them apart. Independence breaks this indeterminacy. The non-cognitive factor is therefore the component of non-cognitive skills that is orthogonal to cognitive ability, and any correlation between the two enters through the cross-loadings of the non-cognitive measurements. With six dedicated cognitive measurements and three non-cognitive ones, the system is over-identified.

5.2 Outcome Equations

The parameters of the measurement system and the factor distribution are held fixed in the second stage. The outcome equations inherit the location, scale and rotation of the factors, so their loadings measure the effect of the cognitive (non-cognitive) factor as defined above.

Conditional independence. The key identifying assumption is that, conditional on the observables and the factors, the measurement errors, the education and occupation preference shocks and the wage shock are mutually independent:

$$\mathbf{e}_i \perp\!\!\!\perp e_i^j \perp\!\!\!\perp \varepsilon_{io} \perp\!\!\!\perp u_i \mid (\mathbf{S}_i, \mathbf{X}_i, \boldsymbol{\theta}_i). \tag{5}$$

All dependence between measurements, education, occupation and wages therefore runs through $\boldsymbol{\theta}_i$ (Carneiro et al., 2003; Heckman, Humphries, et al., 2018). Post-secondary education $\mathbf{P}_i$ and occupational task content $\mathbf{T}_{o_i}$ are endogenous in the wage equation (3) only because they are selected on $\boldsymbol{\theta}_i$. The measurements act as proxies for the factors that are excluded from all outcome equations: the outcome loadings are identified from the covariances between outcomes and measurements, and selection into education levels and occupations is accounted for by the likelihood (11).

The model contains no exclusion restrictions: the lower secondary track and background characteristics enter all equations. The effect of education and task content on wages is therefore identified through the factor structure (5) together with the parametric assumptions on the shocks (Gumbel preference shocks, normal wage shocks) (Carneiro et al., 2003; Heckman, Humphries, et al., 2018). Instruments for the education and occupation choices, such as regional variation in the supply of apprenticeship places, would relax this reliance on functional form.

Normalisations. In the education multinomial logit, we normalise the utility of basic education to zero and the variance of the Gumbel shocks to $\pi^2/6$, so that all coefficients in Equation 1 are expressed relative to the standard deviation of the preference shock. Track-specific choice sets $\mathcal{J}(\mathbf{S}_i)$ require one further normalisation, discussed in Section 4.1. In the occupation conditional logit (2), only utility differences across occupations matter, so indi-

vidual characteristics enter only through their interaction with task content. Identification of the base task valuations requires that the 31×3 matrix of task intensities has full column rank. Each task index is standardised across 3-digit occupations before being averaged to the 31 groups (Section A.2), which fixes the unit of all task coefficients. When interpreting magnitudes, we instead use the standard deviation of each index across first jobs in the estimation sample, about 0.7 (Section 7.4). The Gumbel variance is normalised as in the education choice. Since the rotation of the factors is fixed by the first stage, each element of $\mathbf{\Gamma}$ is the effect of a given ability on the valuation of a given task. Finally, the log wage is continuous, so the scale of its shock is identified, and the wage loadings are expressed in log points per unit of the factor.

6 Estimation

Assuming independence across individuals, the likelihood writes

$$\mathcal{L} = \prod_{i} f(\mathbf{Y_i}, \mathbf{D_i}, \mathbf{M_i} | \mathbf{X_i}) = \prod_{\mathbf{i}} \int_{\mathbf{\Theta}} \mathbf{f}(\mathbf{Y_i}, \mathbf{D_i} | \mathbf{X_i}, \theta) \mathbf{f}(\mathbf{M_i} | \mathbf{X_i}, \theta) \mathbf{f}_\theta(\theta) \mathbf{d}\theta \tag{6}$$

where $\mathbf{Y_i}$ is a vector of outcomes (occupations o or wage w), $\mathbf{D_i}$ denotes the vector of post-secondary schooling decisions, $\mathbf{M_i}$ is a vector a measurements, $\mathbf{X_i}$ is a vector of control variables (gender, household income and a rural dummy, together with the lower secondary track $\mathbf{S_i}$), and θ comprises the cognitive and non-cognitive latent factors, with support Θ.

The likelihood function can be estimated either in a single step or as a two-step procedure. We opt for a two-step procedure, so that the factors are pinned down by the measurement

system alone and are not shaped by the schooling, occupation and wage outcomes. Since the measurements are observed at early ages, before the post-secondary choices, it is natural to treat the measurement system as given when estimating the choices. The two-step procedure also offers computational advantages. In the first stage, we recover parameters from the measurement system. In the second stage, post-secondary education decisions are estimated jointly with outcomes given parameters from the first stage. The full derivation of both likelihoods is given in Appendix C.

6.1 First Stage

The goal of the first stage is to recover parameters from the measurement equations and the distribution of the latent factors. Since both the latent factors and the idiosyncratic shocks are unobserved to the econometrician, we integrate over the factor distribution. Given the normality and mutual independence of the measurement shocks in Equation 4, the first-stage likelihood is

$$\mathcal{L}^1 = \prod_i \int_\Theta \left[\prod_{m=1}^{M} \frac{1}{\sigma_{M_m}} \phi \left(\frac{M_{i,m} - \mu_i^m(\bar{\theta})}{\sigma_{M_m}} \right) \right] f_\theta(\bar{\theta}) d\bar{\theta}, \tag{7}$$

where $\mu_i^m(\bar{\theta}) = \beta_{0,m} + \mathbf{S_i}'\beta_{\mathbf{1,m}} + \mathbf{X_i}'\beta_{\mathbf{2,m}} + \alpha_{\mathbf{m}}^{\mathbf{c}}\bar{\theta}^{\mathbf{c}} + \alpha_{\mathbf{m}}^{\mathbf{nc}}\bar{\theta}^{\mathbf{nc}}$ is the deterministic part of measurement m, σ_{M_m} is the standard error of the measurement shock, and ϕ is the standard normal density function (see Section C.1). Some measurements are missing for part of the sample (see Section 3). Since the measurement shocks are independent conditional on the factors, a missing measurement integrates out of Equation 7: the product over m then runs only over the set $\mathcal{M}_i$ of measurements observed for individual i. This is full-information maximum likelihood, valid if measurements are missing at random given the covariates and the observed measurements. The same restriction applies to the measurement densities in the second stage.

6.2 Second Stage

The second stage recovers the parameters of the schooling, occupation and wage equations, holding the first-stage parameters fixed at their estimates and integrating over the latent abilities. Schooling and occupation utility shocks are Gumbel distributed and the wage shock is normal, so that, conditional on $\theta = \bar{\theta}$, the probability of completing post-secondary level $j \in \mathcal{J}(\mathbf{S}_i)$, the probability of choosing occupation o, and the wage density are

$$\mathcal{P}_i^j(\bar{\theta}) = \frac{\exp\left(\tilde{U}_i^j\right)}{\sum_{k \in \mathcal{J}(\mathbf{S}_i)} \exp\left(\tilde{U}_i^k\right)}, \tag{8}$$

$$\mathcal{P}_{io}(\bar{\theta}) = \frac{\exp\left(\tilde{U}_{io}\right)}{\sum_{o' \in \mathcal{O}} \exp\left(\tilde{U}_{io'}\right)}, \tag{9}$$

$$f(w_i|\bar{\theta}) = \frac{1}{\sigma_w} \phi\left(\frac{w_i - \mu_i^w(\bar{\theta})}{\sigma_w}\right), \tag{10}$$

where $\tilde{U}_i^j = U_i^j - e_i^j$ and $\tilde{U}_{io} = U_{io} - \varepsilon_{io}$ are the deterministic parts of Equation 1 and Equation 2, and $\mu_i^w(\bar{\theta})$ is the deterministic part of the log wage equation Equation 3, evaluated at the task content $\mathbf{T}_{\mathbf{o_i}}$ of the chosen occupation. The second-stage likelihood is then

$$\mathcal{L}^2 = \prod_i \int_\Theta \left[\prod_{m \in \mathcal{M}_i} f(M_{i,m}|\bar{\theta}; \hat{\beta}_{\mathbf{m}}, \hat{\alpha}_{\mathbf{m}}, \hat{\sigma}_{\mathbf{M_m}})\right] \times \prod_{j \in \mathcal{J}(\mathbf{S}_i)} \left[\mathcal{P}_i^j(\bar{\theta})\right]^{\mathbb{1}\{D_i = j\}} \mathcal{P}_{io_i}(\bar{\theta}) \left[f(w_i|\bar{\theta})\right]^{\mathbb{1}\{w_i \text{ observed}\}} f_\theta(\bar{\theta}) d\bar{\theta}, \tag{11}$$

where the measurement densities, evaluated at the first-stage estimates, discipline the distribution of abilities for each individual. The wage density enters only for individuals with an observed wage. The expanded expression is given in Section C.2.

Because the second stage treats the first-stage estimates as known, the inverse Hessian of $\mathcal{L}^2$ understates the sampling variance of the second-stage parameters. We therefore correct the second-stage standard errors for first-stage estimation error following Murphy and Topel

(1985).

6.3 Latent factor integration

We use the Gauss-Hermite quadrature approximation (Ashworth et al., 2021; Heckman, Humphries, et al., 2018) to integrate over the factor distributions. Compared to simulation-based integration, this approach has two advantages: the approximation is exact for polynomials of degree up to $2G-1$ in each factor (with G nodes per dimension), and it requires far fewer evaluations of the integrand. With normal factors $\theta^k \sim \mathcal{N}(0, \sigma_k^2)$, $k \in \{c, nc\}$, each integral over Θ in Equation 7 and Equation 11 is replaced by a weighted sum over a product grid of $Q = G^2$ nodes,

$$\int_{\Theta} h_i(\bar{\theta}) f_\theta(\bar{\theta}) d\bar{\theta} \approx \sum_{q=1}^{Q} \tilde{\omega}_q \, h_i(\bar{\theta}_q),$$
$$\bar{\theta}_q = \left(\sqrt{2}\sigma_c \xi_{q_1}, \sqrt{2}\sigma_{nc} \xi_{q_2}\right), \quad \tilde{\omega}_q = \frac{\omega_{q_1}\omega_{q_2}}{\pi}, \tag{12}$$

where $h_i(\cdot)$ is the individual integrand and $(\omega_{q_k}, \xi_{q_k})$ are the Gauss-Hermite weights and abscissæ. The derivation and the choice of the number of nodes are discussed in Section C.3.

7 Results

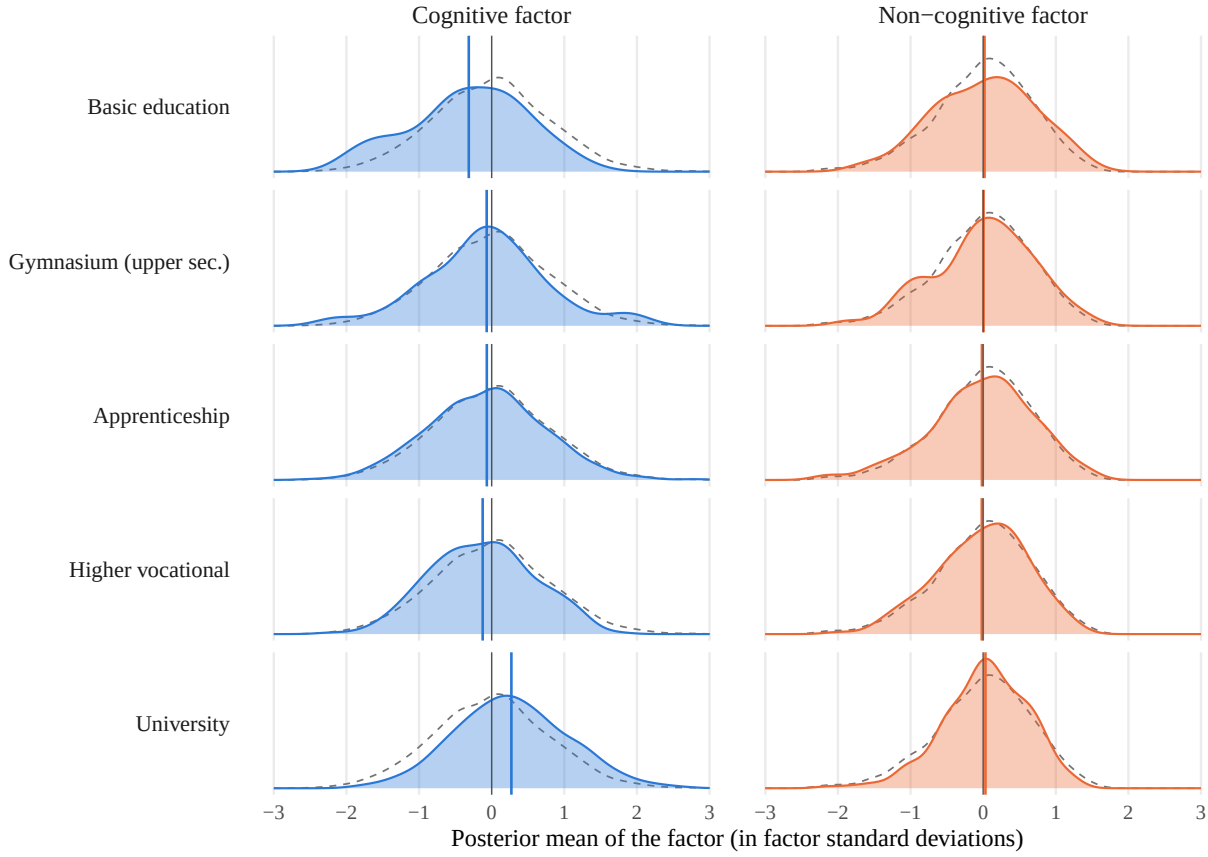
All second-stage estimates in this section come from the joint model of education, occupation and wages (Section 6), estimated on the 1,137 individuals for whom we observe the first occupation. Of these, 1,114 also report an hourly wage; the remaining 23 contribute to the education and occupation parts of the likelihood only. The latent factors are integrated over each individual's posterior distribution from the first stage, evaluated on a 20×20 Gauss–Hermite grid. Standard errors are Murphy–Topel corrected for the first-stage estimation of

the measurement system. The correction is small, and negligible for parameters that do not involve the factors.

7.1 Measurement System

The measurement system is estimated by full-information maximum likelihood on the 1,145 individuals with observed covariates, a superset of the second-stage sample. Table D1 in Section D.1 reports the estimates. Net of the school track, grades are more informative about the cognitive factor than the IST test scores: the factors explain about a third of their residual variance, against well under a fifth for the test scores. Self-esteem is the most informative non-cognitive indicator. The estimated standard deviations of the factors are $\sigma_{\theta^c} = 0.365$ and $\sigma_{\theta^{nc}} = 0.361$. Throughout, we express the effects of ability per standard deviation of the corresponding factor, i.e. the estimated loading multiplied by σ_θ.

Figure 2: Posterior factor means by completed education



Notes: Kernel densities of the first-stage posterior means $\mathbb{E}[\theta_i \mid \mathbf{M}_i, \mathbf{X}_i]$, in units of the estimated factor standard deviation, on the second-stage sample ($N = 1{,}137$). Dashed grey: pooled density. Vertical coloured lines: group means. The non-cognitive panel excludes the 232 individuals with no non-cognitive measurement, whose posterior mean equals the prior mean of zero. Posterior means are shrunk towards zero, so their dispersion understates that of the factors.

Figure 2 shows how the posterior factor means vary with completed education. Because the school track enters every measurement equation, the factors measure ability net of the track, and their means are essentially zero within each track. Sorting on cognitive ability is nonetheless visible. University graduates have a mean cognitive factor about a quarter of a standard deviation above the sample average, and individuals who stop at basic education about a third of a standard deviation below it. The other three degrees lie slightly below the

average. The non-cognitive densities are almost identical across degrees. This anticipates the education choice estimates below, where cognitive ability matters and non-cognitive ability does not.

7.2 Education Choice

Table 3: Post-secondary education choice

	Gymnasium	Apprenticeship	Higher voc.	University
Initial conditions				
Intercept	-1.911** (0.780)	1.343*** (0.270)	-0.626* (0.353)	0.135 (0.388)
Female	-1.017*** (0.389)	-0.433 (0.287)	-0.062 (0.319)	-0.112 (0.347)
HH income (std.)	0.720*** (0.248)	0.582*** (0.207)	0.562** (0.222)	0.764*** (0.230)
Rural	-0.065 (0.442)	0.487 (0.300)	0.462 (0.333)	0.213 (0.365)
Realschule	—	1.307*** (0.296)	1.928*** (0.376)	—
Gymnasium	3.574*** (0.795)	—	2.418*** (0.294)	2.433*** (0.399)
Factor loadings				
Cognitive factor	2.927*** (0.790)	1.638*** (0.550)	1.985*** (0.618)	4.401*** (0.729)
Non-cognitive factor	-0.106 (0.897)	-0.427 (0.603)	-0.431 (0.688)	-0.163 (0.758)

Notes: Base category: basic education. $U_i^j = \beta_0^j + \mathbf{S}_i'\boldsymbol{\beta}_1^j + \mathbf{X}_i'\boldsymbol{\beta}_2^j + \alpha^{c,j}\theta^c + \alpha^{nc,j}\theta^{nc} + e_i^j$. $\mathbf{X}_i$: sex, HH income, rural; $\mathbf{S}_i$: school track (base = Hauptschule). Gymnasium/University are unavailable to Hauptschule students, so within those two columns Realschule and Gymnasium-track are collinear with the intercept; the Realschule coefficient is fixed at 0 there (marked —), making Realschule the implicit reference track for those two columns. Separately, Gymnasium-track's entire choice set is exactly the four free categories, so a common shift across all four Gymnasium-track coefficients is unidentified; the Gymnasium-track coefficient is fixed at 0 for Apprenticeship (marked —) to anchor it. Factor loadings are per unit of the factor (anchored on word analogy and conscientiousness); multiply by $\sigma_{\theta^c} = 0.365$ or $\sigma_{\theta^{nc}} = 0.361$ for a one-SD effect. SEs are Murphy–Topel two-step corrected (Stage-1 uncertainty propagated). * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. $N = 1137$; joint-model log-likelihood -17077.6.

Table 3 reports the post-secondary education choice, with basic education as the base category. Cognitive ability sorts individuals into every post-secondary degree, and most strongly into academic paths. A one standard deviation increase in θ^c raises the log-odds of completing university relative to basic education by 1.6, and of the *Gymnasium* upper-secondary degree by 1.1. The effects on apprenticeship and higher vocational education are less than half as large as on university, but all four are significant at the 1% level. Non-cognitive ability plays no role in the education choice once cognitive ability is accounted for: its loadings are

small and none is significant.

Initial conditions matter a great deal. The reference track differs across columns, because each track has its own choice set.[2] Relative to *Hauptschule*, *Realschule* students are much more likely to complete an apprenticeship or a higher vocational degree rather than stop at basic education. Because basic education is not available to them, the *Gymnasium*-track coefficients compare each degree with an apprenticeship. Compared with *Hauptschule* students, *Gymnasium*-track students are more likely to choose a higher vocational degree over an apprenticeship. Compared with *Realschule* students, their log-odds of the *Gymnasium* degree and of university relative to an apprenticeship are higher by about 4.9 and 3.7, by far the largest effects in the table. Household income raises the probability of every degree relative to basic education, most for university. Women are less likely than men to stop at the *Gymnasium* degree, but not significantly less likely to complete an apprenticeship.

7.3 Occupation Choice

Table 4 reports the occupation choice. Each coefficient is the valuation of a task dimension, so the coefficients describe sorting across the 31 occupations along their task content. Several patterns stand out.

Gender is by far the strongest predictor of sorting. Women value the interpersonal content of an occupation much more than men, and its cognitive and manual content much less, with

[2]The track-specific choice sets $\mathcal{J}(\mathbf{S}_i)$ of Section 4.1 (see also Section 3) imply two normalisations, marked by a dash in Table 3. First, *Hauptschule* students never reach the *Gymnasium* and university outcomes. In these two columns, only *Realschule* and *Gymnasium* students contribute, and the intercept and the two track dummies are not separately identified. We set the *Realschule* coefficient to zero, so *Realschule* is the reference track for these two columns, while *Hauptschule* remains the reference for apprenticeship and higher vocational education. Second, basic education, the base alternative, is not in the choice set of *Gymnasium* students. Adding the same constant to their four track coefficients leaves their choice probabilities unchanged. We set the *Gymnasium*-track coefficient to zero for apprenticeship, which every track can reach. The other *Gymnasium*-track coefficients therefore measure the shift in the utility of each degree *relative to an apprenticeship*.

Table 4: Occupation choice

	Cognitive	Interpersonal	Manual
Factor loadings			
Base task effect	-0.698** (0.314)	0.317 (0.219)	0.033 (0.286)
Cognitive factor	0.631** (0.280)	-0.179 (0.181)	0.066 (0.250)
Non-cognitive factor	0.306 (0.362)	-0.327 (0.238)	0.102 (0.325)
Education × task			
Gymnasium	0.209 (0.488)	-0.262 (0.330)	0.349 (0.449)
Apprenticeship	0.992*** (0.327)	-0.693*** (0.228)	0.851*** (0.300)
Higher voc.	1.273*** (0.376)	-0.712*** (0.255)	0.936*** (0.345)
University	1.086** (0.425)	-0.405 (0.280)	0.450 (0.386)
Background × task			
Female	-1.229*** (0.153)	1.239*** (0.100)	-1.267*** (0.143)
HH income (std.)	-0.008 (0.081)	-0.021 (0.051)	-0.143* (0.073)
Rural	0.145 (0.157)	-0.120 (0.102)	0.085 (0.144)
Realschule	0.353 (0.222)	0.001 (0.157)	-0.135 (0.207)
Gymnasium	0.492* (0.278)	0.059 (0.188)	-0.493* (0.256)

Notes: Conditional logit over 31 occupations. Task dimensions: Yamaguchi directed-PCA indices (three linearly independent factors). $U_{io} = \sum_{k\in\{c,ip,m\}} T_o^k(\delta^k + \Gamma_{k,c}\theta^c + \Gamma_{k,nc}\theta^{nc} + \mathbf{P}_i'\boldsymbol{\phi}^k + \mathbf{X}_i'\boldsymbol{\psi}^k)$. $\boldsymbol{\Gamma}$ is a freely estimated 3×2 matrix — both factors enter all task dimensions. $\mathbf{P}_i$: post-secondary education dummies (base = Basic ed.). $\mathbf{X}_i$: sex, HH income, rural, school track (base = Hauptschule). Factor loadings are per unit of the factor (anchored on word analogy and conscientiousness); multiply by $\sigma_{\theta^c} = 0.365$ or $\sigma_{\theta^{nc}} = 0.361$ for a one-SD effect. SEs are Murphy–Topel two-step corrected (Stage-1 uncertainty propagated). * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. $N = 1137$; joint-model log-likelihood -17077.6.

coefficients of similar magnitude on all three dimensions. Because occupations rich in cognitive tasks are also rich in interpersonal tasks, these coefficients should be read jointly: women sort towards interpersonal occupations, given their cognitive and manual content.

Post-secondary education shapes sorting along all three dimensions. Relative to basic education, apprentices and higher vocational graduates value cognitive and manual content more and interpersonal content less, and the two vocational degrees have very similar coefficients. University graduates also value cognitive content more; their lower valuation of interpersonal content and higher valuation of manual content are smaller and not significant. Holding the other two dimensions fixed, any post-secondary degree therefore moves individuals away from interpersonal occupations, and vocational degrees move them towards manual ones.

Individuals with only basic education are a small group (68 individuals), so these contrasts remain imprecise.

Latent abilities add little once education and background are controlled for. A one standard deviation increase in θ^c raises the valuation of cognitive tasks by 0.23, about a fifth of the effect of a post-secondary degree. The other entries of $\mathbf{\Gamma}$ are not significant. Ability sorting into occupations therefore operates mostly through the education choice, where the cognitive factor is strongly significant.

7.4 Wages

Table 5 reports the wage equation. The outcome is the log hourly wage in the first job. Conditional on the task content of the occupation, background and latent abilities, returns to post-secondary education rise with the level of the degree. Relative to basic education, an apprenticeship raises hourly wages by about 11%, a higher vocational degree by 28% and a university degree by 45%; the last two are strongly significant, the apprenticeship premium only marginally so. University graduates earn about 31% more than apprentices, a precisely estimated gap. An upper-secondary *Gymnasium* degree without further education earns about 18% less than basic education. The lower-secondary track has an effect of its own: whatever their final degree, individuals who attended *Gymnasium* earn about 17% more than those who attended *Hauptschule*, and those who attended *Realschule* about 8% more. There is no gender gap in first-job hourly wages once occupational sorting is accounted for.

Occupations are paid for their task content. Within the estimation sample, the standard deviation of each task index across first jobs is about 0.7. A one standard deviation move towards a more cognitive occupation raises wages by about 10%, and towards a more interpersonal one lowers them by about 7%. The return to manual content is small and not significant. These are partial effects. Cognitive and manual content are strongly negatively

Table 5: Wage equation

	Log wages
Intercept	2.110*** (0.062)
Education (base = Basic ed.)	
Gymnasium	-0.202** (0.093)
Apprenticeship	0.105* (0.063)
Higher voc.	0.249*** (0.070)
University	0.375*** (0.077)
Background	
Female	0.020 (0.029)
HH income (std.)	0.032** (0.014)
Rural	-0.030 (0.028)
Realschule	0.076* (0.043)
Gymnasium	0.156*** (0.052)
Task returns	
Cognitive	0.136*** (0.039)
Interpersonal	-0.110*** (0.025)
Manual	0.030 (0.037)
Factor loadings	
Cognitive factor	0.052 (0.052)
Non-cognitive factor	0.066 (0.067)
Residual std. dev.	0.451

Notes: Outcome: log real gross hourly wage in the first job. $\log w_i = \pi_0 + \mathbf{P}_i'\boldsymbol{\pi}_P + \mathbf{X}_i'\boldsymbol{\pi}_X + \boldsymbol{\rho}'\mathbf{T}_{o_i} + \alpha_w^c\theta^c + \alpha_w^{nc}\theta^{nc} + u_i$, $u_i \sim N(0, \sigma_w^2)$. $\mathbf{P}_i$: post-secondary education dummies (base = Basic ed.). $\mathbf{X}_i$: sex, HH income, rural, school track (base = Hauptschule). $\mathbf{T}_{o_i}$: task intensities of chosen occupation. Factor loadings are per unit of the factor (anchored on word analogy and conscientiousness); multiply by $\sigma_{\theta^c} = 0.365$ or $\sigma_{\theta^{nc}} = 0.361$ for a one-SD effect. SEs are Murphy–Topel two-step corrected (Stage-1 uncertainty propagated). * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. 1114 individuals have an observed wage. $N = 1137$; joint-model log-likelihood -17077.6.

correlated across jobs, so moving from a manual to a cognitive occupation trades a gain on one dimension against a loss on the other.

Finally, the latent factors have small and imprecise direct effects on wages once education and occupation are controlled for. A one standard deviation increase raises wages by about 2% for either cognitive or non-cognitive ability, and neither effect is significant. Cognitive ability affects wages mainly indirectly, through the education choice.

7.5 Decomposition: Education and Occupation Channels

Education affects wages through two channels. First, it shifts wages directly within an occupation ($\boldsymbol{\pi}_P$). Second, it changes the occupations individuals sort into ($\boldsymbol{\phi}^k$), whose task content is priced in the wage equation ($\boldsymbol{\rho}$). Symmetrically, the raw wage gap between occupations combines the price of their task content with the composition of their workforce. We quantify each channel by shutting down the other one in the estimated model.

Shutting down the occupation channel. For each education level j we compute the average effect, over the estimation sample and the posterior distribution of the factors, of assigning everyone to j rather than to basic education:

$$\mathrm{TE}_j = \underbrace{\pi_{P,j}}_{\text{direct}} + \underbrace{\boldsymbol{\rho}' \left(\mathbb{E}[\mathbf{T}_o \mid P = j] - \mathbb{E}[\mathbf{T}_o \mid P = \text{Basic}] \right)}_{\text{via occupational sorting}},$$

where $\mathbb{E}[\mathbf{T}_o \mid P = j]$ is the expected task content of the chosen occupation when education is set to j, keeping each individual's characteristics and abilities fixed. The direct effect is the wage gain if the distribution of occupations is held at its basic-education level, i.e. with the occupation channel shut down.

Table 6 shows that the occupation channel is small but precisely estimated. For each degree, sorting into occupations with better-paid task content raises wages by about 3% relative to basic education. This is 8% of the total university premium, 12% of the higher vocational premium and 20% of the apprenticeship premium. Since post-secondary degrees move individuals away from interpersonal occupations, which pay less, every degree gains a little from occupational sorting. The channel plays no role in the comparison between university and apprenticeship: it accounts for only 2% of their gap, and is not significant. Overall, returns to vocational and academic degrees arise mostly within occupations, not from access

Table 6: Education and log wages: direct and occupation channels

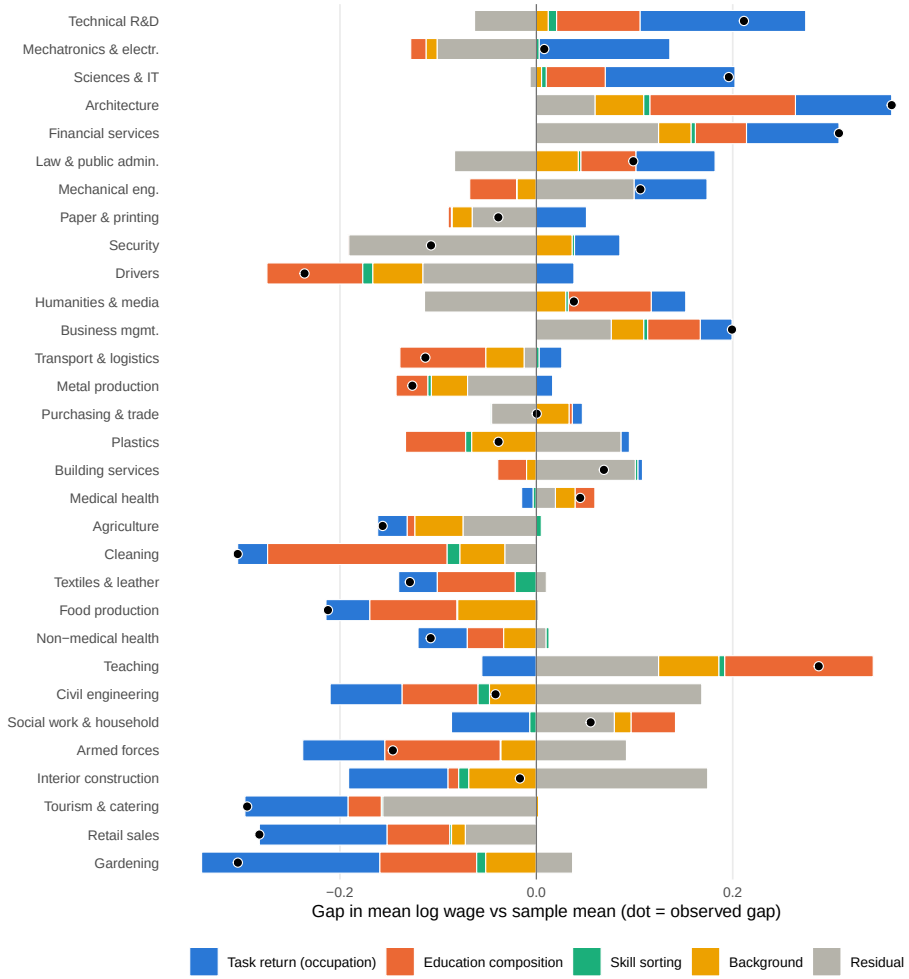
Contrast	Direct	Via occupation	Total	Occ. share
Gymnasium vs Basic ed.	-0.202 (0.093)	0.004 (0.016)	-0.198 (0.094)	–
Apprenticeship vs Basic ed.	0.105 (0.063)	0.027 (0.013)	0.132 (0.064)	20%
Higher voc. vs Basic ed.	0.249 (0.070)	0.034 (0.014)	0.283 (0.071)	12%
University vs Basic ed.	0.375 (0.077)	0.032 (0.016)	0.407 (0.079)	8%
University vs Apprenticeship	0.269 (0.047)	0.006 (0.011)	0.275 (0.047)	2%

Notes: Model-implied effect on log hourly wages of assigning everyone to the row degree j rather than the comparison degree k, holding background X and latent abilities θ fixed. *Direct*: wage effect of the degree within a given occupation (π_P in Table 5), i.e. holding the distribution of chosen occupations at the one under the comparison degree. *Via occupation*: wage value of the change in the task content of the chosen occupation that the degree induces, $\rho'(E[T \mid P = j] - E[T \mid P = k])$. *Total*: sum of the two. *Occ. share*: via occupation / total, not reported when the total is negative. Averages over the estimation sample, integrated over individual posterior factor weights. Delta-method SEs (Murphy–Topel corrected) in parentheses. $N = 1137$.

to different occupations.

Shutting down the education channel. Since the wage equation has no interaction between education and tasks, the wage return to occupation o is the same for every education level once education is held fixed, $\boldsymbol{\rho}'\mathbf{T}_o$. Across the 31 occupations, the same worker would earn about 22% more in the occupation at the 90th percentile of this return than in the one at the 10th percentile. We then decompose the gap between an occupation's mean log wage and the sample mean into (i) its task return $\boldsymbol{\rho}'\mathbf{T}_o$, (ii) the education composition of its workers $\mathbb{E}[\mathbf{P}'\boldsymbol{\pi}_P \mid o]$, (iii) skill sorting $\mathbb{E}[\alpha_w^c\theta^c + \alpha_w^{nc}\theta^{nc} \mid o]$, (iv) background $\mathbb{E}[\mathbf{X}'\boldsymbol{\pi}_X \mid o]$, and (v) a residual.

Figure 3: Decomposition of occupation wage gaps



Notes: Gap between the mean log hourly wage of each occupation and the sample mean (dot), decomposed into the occupation's task return $\boldsymbol{\rho}'\mathbf{T}_o$ (education channel shut), education composition, skill sorting, background and residual. Occupations (KldB 2010, 2-digit) sorted by task return. See Appendix Table D2 for the numbers.

Figure 3 shows the result. Employment-weighted, the task return and the education composition of the workforce each account for slightly less than 30% of the between-occupation variance in mean log wages. Background accounts for about a tenth, skill sorting for almost nothing, and the model leaves 30% unexplained. The occupations that pay best for a given worker are technical R&D, mechatronics and electrical, and natural sciences and IT. Gardening, retail sales, tourism and catering, and interior construction pay least. Some occupations

owe their high average wage mostly to the education of their workforce. In teaching and architecture, for instance, education composition is larger than the task return. Thus slightly less than 30% of the wage differences across first jobs reflects what the occupation pays for a given worker. Who sorts into each occupation, mainly in terms of education and background, explains about 40%.

Re-estimation with each channel removed. The decomposition above uses the estimates of the full model. As a check, we re-estimate the joint model twice: once without task returns in the wage equation ($\boldsymbol{\rho} = \mathbf{0}$), and once without education ($\boldsymbol{\pi}_P = \mathbf{0}$). All three likelihood blocks are kept in each case (Appendix Table D3). Unlike the decomposition, which holds the estimates fixed, the remaining parameters now adjust to absorb the omitted channel. Without task returns, the education coefficients measure the total effect of education, and they are almost identical to the totals in Table 6. Without education, the interpersonal and manual returns barely change, and the cognitive return rises only from 0.14 to 0.17, so the task returns do not mainly proxy for the education of the workforce. The wage loading on cognitive ability, by contrast, more than triples and becomes significant: a one standard deviation increase in θ^c now raises wages by about 6%, against 2% in the full model. Once education is removed, cognitive ability picks up its effect, consistent with cognitive ability affecting wages mainly through the education choice.

8 Conclusion and Next Steps

To summarise, cognitive ability sorts individuals strongly into post-secondary education, especially academic degrees, while non-cognitive ability plays no role. Occupational sorting along task content is driven mainly by gender and education, and latent abilities add little once these are controlled for. Wage returns rise with the level of the degree and arise mostly

within occupations: the occupation channel accounts for 8 to 20% of the education premia. Conversely, what an occupation pays for a given worker explains slightly less than 30% of the wage differences across first jobs, and the composition of its workforce about 40%.

We are currently bootstrapping the standard errors of the full two-step estimator, re-estimating both the measurement system and the joint model on each bootstrap sample so that the uncertainty in the latent factor distribution carries through to the Stage-2 parameters. These bootstrap standard errors will be reported in the next draft. Beyond inference, our model estimates will be used to simulate how a change in the work environment or the adoption of a counterfactual policy that would alter the trade-offs faced by agents between working conditions, study conditions and labor market wage would affect enrollment rates across education programs and occupational outcomes. It will also be used to calculate heterogeneous treatment effects by computing counterfactual wages for students if they had chosen a different educational path given their unobserved abilities.

References

Altonji, J. G., P. Arcidiacono, and A. Maurel (2016). "The Analysis of Field Choice in College and Graduate School. Determinants and Wage Effects". In: *Handbook of the Economics of Education*. 1st ed. Vol. 5. Elsevier B.V., pp. 305–396.

Amthauer, R. et al. (2001). *I-S-T 2000 R Intelligenz-Struktur-Test 2000 R.* Tech. rep.

Anderson, T. W. and H. Rubin (1956). "Statistical Inference in Factor Analysis". *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability, Volume 5: Contributions to Econometrics, Industrial Research, and Psychometry* 3.5, pp. 111–151.

Arcidiacono, P. (2004). "Ability sorting and the returns to college major". *Journal of Econometrics* 121 (1-2), pp. 343–375.

Ashworth, J., V. J. Hotz, and A. Maurel (2021). "Changes across cohorts in wage returns to schooling and early work experiences". *Journal of Labor Economics* 39 (4), pp. 931–964.

Beffy, M., D. Fougère, and A. Maurel (2012). "Choosing the field of study in postsecondary education: Do expected earnings matter?" *Review of Economics and Statistics* 94 (1), pp. 334–347.

Belzil, C. and F. Poinas (2018). "Estimating a Model of Qualitative and Quantitative Education Choices in France".

Biewen, M. and M. Tapalaga (2017). "Life-cycle educational choices in a system with early tracking and 'second chance' options". *Economics of Education Review* 56, pp. 80–94.

Bundesinstitut für Berufsbildung (2019). *Datenreport zum Berufsbildungsbericht*. Tech. rep.

Carneiro, P., K. T. Hansen, and J. J. Heckman (2003). "Estimating distributions of treatment effects with an application to the returns to schooling and measurement of the effects of uncertainty on college choice". *International Economic Review* 44 (2), pp. 361–422.

Castex, G. and E. K. Dechter (2014). "The changing roles of education and ability in wage determination". *Journal of Labor Economics* 32 (4), pp. 685–710.

Cunha, F. et al. (2010). "Estimating the Technology of Cognitive and Noncognitive Skill Formation". *Econometrica* 78 (3), pp. 883–931.

De Groote, O. and K. Declercq (2021). "Tracking and specialization of high schools: Heterogeneous effects of school choice". *Journal of Applied Econometrics* 36 (7), pp. 898–916.

Dengler, K., B. Matthes, and W. Paulus (2014). "Occupational Tasks in the German Labour Market - An alternative measurement on the basis of an expert database".

Dustmann, C., P. A. Puhani, and U. Schönberg (2017). "The Long-term Effects of Early Track Choice". *Economic Journal* 127 (603), pp. 1348–1380.

Eckardt, D. (2026). "Training Specificity and Occupational Mobility : Evidence from German Apprenticeships". *Econometrica*.

European Commission (2018). *Germany: Policy developments on apprenticeship*. Tech. rep.

Eurostat (2025). *Eurostat*.

Gathmann, C. and U. Schönberg (2010). "How general is human capital? A task-based approach". *Journal of Labor Economics* 28 (1), pp. 1–49.

Heckman, J., J. E. Humphries, and G. Veramendi (2018). "Returns to education: The causal effects of education on earnings, health, and smoking". *Journal of Political Economy* 126, S197–S246.

Heckman, J. J. and Y. Rubinstein (2001). "The benefits of skill: The importance of noncognitive skills: Lessons from the GED testing program". *American Economic Review* 91 (2), pp. 145–154.

Heckman, J. J., J. Stixrud, and S. Urzua (2006). "The effects of cognitive and noncognitive abilities on labor market outcomes and social behavior". *Journal of Labor Economics* 24 (3), pp. 411–482.

Heckman, J. J. and E. Vytlacil (2001). "Policy-relevant treatment effects". *American Economic Review* 91 (2), pp. 107–111.

Innengruppe Bildungsberichterstattung (2024). *Bildung in Deutschland*. Tech. rep.

Kinsler, J. and R. Pavan (2015). "The specificity of general human capital: Evidence from college major choice". *Journal of Labor Economics* 33 (4), pp. 933–972.

Lemieux, T. (2014). "Occupations, fields of study and returns to education". *Canadian Journal of Economics* 47 (4), pp. 1047–1077.

Murphy, K. M. and R. H. Topel (1985). "Estimation and Inference in Two-Step Econometric Models". *Journal of Business & Economic Statistics* 3 (4), p. 370.

Rotter, J. B. (1966). "Generalized Expectancies for Internal versus External Control of Reinforcement". *Psychological Monographs: General and Applied* 80 (1).

Spitz-Oener, A. (2006). "Technical change, job tasks, and rising educational demands: Looking outside the wage structure". *Journal of Labor Economics* 24 (2), pp. 235–270.

Yamaguchi, S. (2012). "Tasks and heterogeneous human capital". *Journal of Labor Economics* 30 (1), pp. 1–53.

A Data

A.1 Measurements

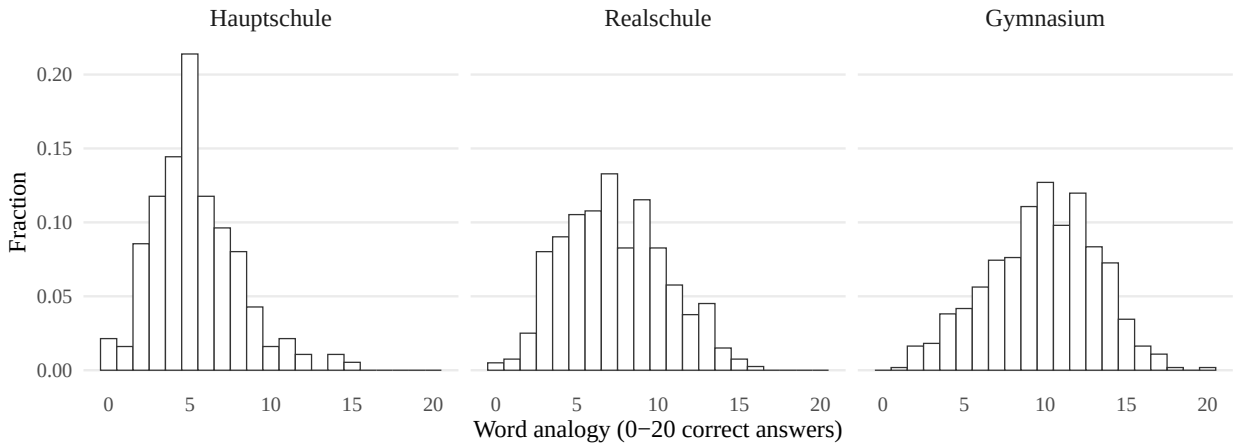
An attractive GSOEP feature is its wide range of both cognitive and non-cognitive measures. A brief overview of the ones that will be used to estimate our model:

$$\begin{cases} \text{cognitive:} & \text{IST 2000R test (Word Analogy, Arithmetic, Matrices)} \\ & \text{German, Foreign Language and Maths school grades} \\ \text{noncognitive:} & \text{Locus of control, Conscientiousness, Self-Esteem} \end{cases}$$

Word Analogy, Arithmetic and Matrices come directly from the IST 2000R test, in which respondents answer 20 question in each category. School grades in German, Foreign Language and Mathematics are self-reported at the age of 17.

To illustrate their informativeness, we plot the number of correct answers (out of 20) in "word analogy", part of the IST 2000R test.

Figure A1: Word Analogy by High School track



As can be seen in Figure A1, there seem to descriptive evidence of ability sorting. Students enrolled in the academic *Gymnasium* track outperform individuals in *Realschule* and *Hauptschule*.

Regarding non-cognitive measures, we include Locus of control, which describes the extent to which individuals believe they can determine events in their own lives (Rotter, 1966). Questions include "How much your CV depends on you", "Didn't get what I deserve", "What you achieve depends on luck", "Others' actions determine my life", "Doubt about ability when faced with difficulties", "Possibilities determined by social circumstances", and "Control over things in life" in a scale from one to seven. The average of the answers, ranked from one to seven, is then computed. Conscientiousness is part of the *Big five personality traits*. Questions include "does a thorough job", "tends to be lazy", and "does things effectively and efficiently". Similarly, the average of the measures ranging from one to seven is computed. Finally, self-esteem is a self-reported number from one to seven. All three are taken from the youth questionnaire at age 17, the same year as the cognitive test, so that every measurement predates the post-secondary education choice. They are missing for part of the sample, self-esteem in particular, which the first stage handles by full-information maximum likelihood (Section 6.1). The correlations in Table A1 are computed on pairwise available observations.

Table A1: Measurement correlation

	Word Analogy	Arithmetic	Matrices	German	Foreign Language	Maths	Locus of Control	Conscientiousness	Self-esteem
Word Analogy	1	.3985984	.4833507	.2303071	.1971025	.2655316	.050182	-.0789991	.0159174
Arithmetic	.3985984	1	.4146279	.1188832	.1178771	.2299733	.0735897	.0087017	.032607
Matrices	.4833507	.4146279	1	.1028171	.0569797	.1924105	.080073	-.0566552	-.0148488
German	.2303071	.1188832	.1028171	1	.4936754	.3702006	.114896	.182184	.0865771
Foreign Language	.1971025	.1178771	.0569797	.4936754	1	.3590084	.109172	.1601692	.0637358
Maths	.2655316	.2299733	.1924105	.3702006	.3590084	1	.1144541	.1578353	.0996053
Locus of Control	.050182	.0735897	.080073	.114896	.109172	.1144541	1	.1645286	.290158
Conscientiousness	-.0789991	.0087017	-.0566552	.182184	.1601692	.1578353	.1645286	1	.2131509
Self-esteem	.0159174	.032607	-.0148488	.0865771	.0637358	.0996053	.290158	.2131509	1

A.2 Task Classification and Principal Component Analysis

Task classification. Each of the 18 activities of the BIBB/BAuA 2018 survey is assigned to exactly one task category (Table A3). The assignment follows the five task types of Dengler et al. (2014), themselves based on Spitz-Oener (2006), collapsed into three dimensions:

- analytical non-routine and cognitive routine activities are classified as cognitive;
- interactive non-routine activities are classified as interpersonal;
- manual routine and non-routine activities are classified as manual.

This differs from an intuitive hand classification for two items. Organising and planning the work of others (F310) is interactive in Dengler et al. (2014), hence interpersonal rather than cognitive. Nursing, caring and healing (F316) is manual non-routine, hence manual rather than interpersonal.

Directed PCA. Answers are recoded so that higher values indicate more frequent performance (1 = never, 2 = sometimes, 3 = frequently). They are then averaged within each 3-digit occupation of the KldB 2010 classification. Following Yamaguchi (2012), we run a separate principal component analysis within each task category on these occupation means, using the correlation matrix. We retain the first component, orient it so that it correlates positively with a reference activity (developing/researching for cognitive, manufacturing for manual, teaching for interpersonal), and standardise it across 3-digit occupations. Finally, scores are averaged to the 2-digit level. Natural sciences, geography and IT (KldB 41–43) are merged into one group, and so are humanities, arts and media (91–94). Mining (21) is dropped. This gives the 31 occupations of the choice model. As a robustness check, we also compute the task shares of Spitz-Oener (2006): for each worker, the share of performed activities in each category, averaged by occupation.

Table A2 reports the loadings of the first component. In each category the first component captures 41–54% of the variance, and all loadings but one share the same sign. It can therefore be read as the overall intensity of the category. The exception is measuring, testing and quality control (F304), which loads negatively on the cognitive index. Across occupations, this activity is typical of production jobs rather than of jobs that rely on computers, information gathering and research, and the cognitive index mostly captures the latter. The three indices are not orthogonal: in the estimation sample, the correlation between the cognitive and manual indices is −0.85, between cognitive and interpersonal 0.50, and between manual and interpersonal −0.52. This is why the coefficients on task content in the occupation and wage equations should be read as partial effects.

Table A2: Directed PCA: first-component loadings

Activity	Item	PC1 loading
Cognitive (157 occupations; PC1 explains 54% of variance, eigenvalue 2.72)		
Measuring, testing, quality control	F304	−0.170
Developing, researching, constructing	F311[a]	0.322
Gathering information, documenting	F313	0.557
Working with computers	F318	0.527
Internet, e-mail	F319	0.528
Manual (159 occupations; PC1 explains 41% of variance, eigenvalue 3.25)		
Manufacturing, producing goods	F303[a]	0.379
Monitoring, controlling machines	F305	0.438
Repairing, refurbishing	F306	0.454
Transporting, storing, shipping	F308	0.411
Entertaining, accommodating, food	F315	0.092
Nursing, caring, healing	F316	0.109
Protecting, guarding, patrolling	F317	0.251
Cleaning, waste, recycling	F320	0.455
Interpersonal (159 occupations; PC1 explains 44% of variance, eigenvalue 2.20)		
Purchasing, procuring, selling	F307	0.314
Advertising, marketing, PR	F309	0.457
Organising, planning work of others	F310	0.453
Training, teaching, educating	F312[a]	0.384
Providing advice and information	F314	0.583

Notes: BIBB/BAuA 2018. One PCA per task category on the correlation matrix of 3-digit occupation means of the recoded activity frequencies (unweighted). [a] Reference activity used to sign the component.

Table A4 reports the indices by 2-digit occupation, and Table A5 and Figure A2 the resulting task content of first jobs by completed education. University graduates work in the most cognitive and interpersonal and least manual occupations. Apprentices and individuals with basic education work in the least cognitive and most manual ones. Higher vocational graduates sit in between: they are close to the mean on cognitive and manual content and above it on interpersonal content. Individuals with only basic education or an upper-secondary *Gymnasium* degree are fewer (68 and 64).

Table A3: Task categorization

Variable	**Description**	**Cognitive**	**Manual**	**Interpersonal**
F303	Manufacturing, producing goods and commodities		Yes	
F304	Measuring, testing, quality control	Yes		
F305	Monitoring, control of machines, plants, technical processes		Yes	
F306	Repairing, refurbishing		Yes	
F307	Purchasing, procuring, selling			Yes
F308	Transporting, storing, shipping		Yes	
F309	Advertising, marketing, public relations			Yes
F310	Organising, planning and preparing work processes. Here we are not talking about your own work processes			Yes
F311	Developing, researching, constructing	Yes		
F312	Training, instructing, teaching, educating			Yes
F313	Gathering information, investigating, documenting	Yes		
F314	Providing advice and information			Yes
F315	Entertaining, accommodating, preparing food		Yes	
F316	Nursing, caring, healing		Yes	
F317	Protecting, guarding, patrolling, directing traffic		Yes	
F318	Working with computers	Yes		
F319	Using the Internet or editing e-mails	Yes		
F320	Cleaning, removing waste, recycling		Yes	

Table A4: Task indices by 2-digit occupation

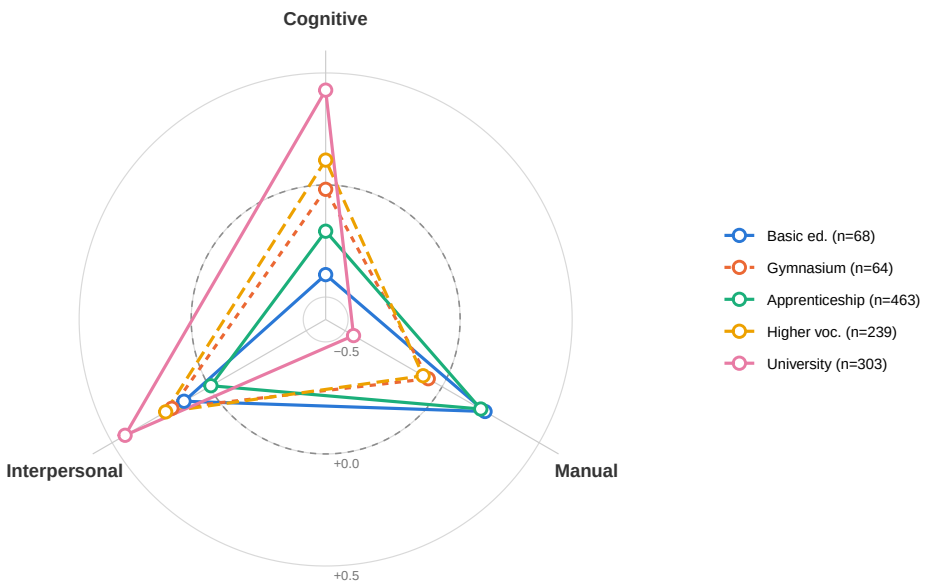
	Cog	Manual	Interpersonal	Cog_SO	Manual_SO	Interpersonal_SO
Armed forces	.5947463	.2644741	1.71685	-.8445041	.3448525	.4483793
Agriculture, animal care, forestry	-.5827743	1.225741	.0494273	-.6605828	.7630148	-.5248861
Gardening and floristry	-1.153006	1.041559	.6703947	-.723684	.5134186	-.0025758
Mining, raw-material extraction and processing	-1.380737	.7642148	-1.158465	.5104772	.5190414	-1.562433
Plastics production and processing	-1.310437	.9345994	-1.274065	-.5763084	1.181041	-1.372773
Paper and printing, media technology	-.0338399	.4758288	-.2161743	-.3494841	.5719758	-.576448
Metal production and processing, metal	-1.151951	1.267819	-1.060402	-.6614712	1.242061	-1.374142
Mechanical and automotive engineering	-.2235858	.5175909	-.6486923	.0193554	.4826242	-.8810301
Mechatronics, energy, and electrical	.2653208	.4122751	-.6101926	.258267	.2608942	-.7874577
Technical research, development, and construction	1.09358	-.7221414	-.2258012	1.18516	-.7524448	-.1526479
Textile and leather	-1.413149	.9269511	-.9677999	-.8463079	1.047119	-.7959727
Food production and processing	-1.297415	1.52112	-.6151203	-1.003416	1.223747	-.9122135
Construction planning, architecture	.7162814	-.9652022	-.1147558	1.371891	-1.007771	.0660908
Civil and structural engineering	-1.265955	.9215273	-.4808408	-.424273	.4253016	-.2221605
Interior construction	-1.418365	.7910947	-.4507801	-.5409421	.623987	-.4283396
Building and supply engineering	-.6764361	1.097858	-.4159501	-.6129792	.7788858	-.6128403
Mathematics, biology and physics	.6715639	-.1562858	-.3888612	.6870949	-.2492562	-.4203992
Geology and environmental protection	.8459695	-.5240907	.4594172	.5938303	-.6285815	.3700774
IT and communications technology	1.234367	-.6496559	-.0227864	1.068139	-.7478859	-.0137843
Transport and logistics	-.5271634	.2719517	-.6255943	-.2312256	.3818634	-.3874832
Transport equipment	-.6560715	.6755816	-.8100165	-.5174192	.9840872	-1.097106
Security, protection, and monitoring	.4468477	-.5361046	.1407489	.1838077	-.3846994	.4520654
Cleaning occupations	-2.117317	-.1251062	-2.200339	-.7143477	.8707094	-.6485398
Purchasing, sales, and trade	.7059181	-.8967899	.6856826	.3398433	-.9942746	1.338193
Sales (retail, etc.)	-.5174133	-.0245945	.6936911	-.697424	.0276519	.8267485
Tourism, hotel, and catering	-.023418	-.2331717	1.016738	-.5645168	-.1215184	.9246392
Business management and organisation	.9004255	-1.128153	.6611362	.6344983	-1.134559	1.217184
Financial services	.7204848	-1.366793	-.1848303	1.303706	-1.453369	.9427154
Legal and public administration	.8055293	-1.076567	.1210661	.8219885	-1.109348	.9369783
Medical health professions	.3113973	.0597756	.6624691	-.3748935	.2770768	-.0210526
Non-medical health	-.1360717	.1276637	.4795625	-.5415492	.3036278	.1411041
Education, social, and household	-.2137811	.1110395	.6474899	-.9472632	.5001056	.3018304
Teaching and training	.7659928	-.8678919	1.362548	.193478	-.7596105	1.105439
Humanities, and social science	1.229469	-1.490029	.3750337	1.798861	-1.688607	.7384722
Advertising and marketing	1.020075	-1.171398	.6930135	.6774498	-1.270413	1.404405
Product design and arts	-.0132131	.2698957	.0044264	-.0739777	.1212562	-.122344
Performing and entertainment	.2006981	-.3986969	.2516969	-.1334923	-.1687128	.4671296
Observations	143					

Notes: Cog, Manual, Interpersonal: directed-PCA indices (first component, standardised across 3-digit occupations, averaged within 2-digit KldB 2010 groups). _SO: Spitz-Oener (2006) task shares, standardised. In the choice model, groups 41–43 and 91–94 are merged and group 21 is dropped.

Education	N	Cognitive	Interpersonal	Manual
Basic ed.	68	-0.40 (0.74)	0.13 (0.92)	0.22 (0.58)
Gymnasium	64	-0.02 (0.64)	0.19 (0.76)	-0.07 (0.63)
Apprenticeship	463	-0.21 (0.70)	-0.01 (0.71)	0.20 (0.75)
Higher voc.	239	0.11 (0.62)	0.22 (0.59)	-0.10 (0.66)
University	303	0.42 (0.53)	0.43 (0.62)	-0.46 (0.62)

Table A5: Task indices of occupation by completed education: mean (SD). Estimation sample.

Figure A2: Mean task profile of first jobs by completed education



Notes: Each vertex is the mean directed-PCA index of the first-job occupation, by completed education. The radius is linear in the index, from -0.6 at the centre to 0.6 on the outer circle. Rings mark -0.5, 0 (dashed) and $+0.5$ standard deviations across 3-digit occupations.

B Identification of the Measurement System

This appendix gives the formal identification argument for the measurement system of Section 5.1. It follows Anderson and Rubin (1956), Carneiro et al. (2003) and Heckman, Stixrud, et al. (2006).

Covariance structure. Stack the nine measurement equations (4) as

$$\mathbf{M}_i = \mathbf{B}\mathbf{Z}_i + \mathbf{A}\boldsymbol{\theta}_i + \mathbf{e}_i, \tag{13}$$

where $\mathbf{Z}_i = (1, \mathbf{S}_i', \mathbf{X}_i')'$, $\mathbf{B}$ collects the intercepts and covariate coefficients, $\mathbf{A}$ is the 9×2 loading matrix, and $\mathbf{e}_i = (e_i^1, \ldots, e_i^9)'$ has variance-covariance matrix $\boldsymbol{\Psi} = \text{diag}(\sigma^2_{M_1}, \ldots, \sigma^2_{M_9})$.

Under (A3), $\boldsymbol{\Sigma}_\theta = \mathrm{Var}(\boldsymbol{\theta}_i) = \mathrm{diag}(\sigma^2_{\theta^c}, \sigma^2_{\theta^{nc}})$. Under (A1) and (A2), $\mathbf{B}$ is identified from the conditional mean $\mathbb{E}[\mathbf{M}_i|\mathbf{S}_i, \mathbf{X}_i] = \mathbf{B}\mathbf{Z}_i + \mathbf{A}\mathbb{E}[\boldsymbol{\theta}_i]$ once the location of the factors is fixed. For any vector $\mathbf{c}$, the factors $\boldsymbol{\theta}_i + \mathbf{c}$ with intercepts $\beta_0^m - \mathbf{A}_m\mathbf{c}$ generate the same distribution of $\mathbf{M}_i$, hence the normalisation $\mathbb{E}[\boldsymbol{\theta}_i] = \mathbf{0}$. The residualised measurements $\tilde{\mathbf{M}}_i = \mathbf{M}_i - \mathbf{B}\mathbf{Z}_i = \mathbf{A}\boldsymbol{\theta}_i + \mathbf{e}_i$ have covariance matrix

$$\boldsymbol{\Sigma}_{\tilde{M}} = \mathbf{A}\boldsymbol{\Sigma}_\theta\mathbf{A}' + \boldsymbol{\Psi}. \tag{14}$$

Its off-diagonal elements depend only on $\mathbf{A}$ and $\boldsymbol{\Sigma}_\theta$, and identification of the factor structure comes from them.

Scale and rotation. For any $\kappa_f > 0$, the loadings α^f_m/κ_f and the factor $\kappa_f\theta^f$ are observationally equivalent, which motivates the normalisations $\alpha^c_1 = 1$ (word analogy) and $\alpha^{nc}_7 = 1$ (conscientiousness). These require that the anchoring measure does load on the factor; if its true loading were close to zero, the normalisation would be fragile. For any invertible 2×2 matrix $\mathbf{R}$, the pair $(\mathbf{A}\mathbf{R}^{-1}, \mathbf{R}\boldsymbol{\Sigma}_\theta\mathbf{R}')$ produces the same $\boldsymbol{\Sigma}_{\tilde{M}}$. With K factors, $\mathbf{R}$ has K^2 free elements, of which the scale normalisations remove K; the remaining $K(K-1)$, two in our case, are removed by the dedicated cognitive measurements and the independence of

the factors. The loading matrix is

$$\mathbf{A} = \begin{pmatrix} 1 & 0 \\ \alpha_2^c & 0 \\ \alpha_3^c & 0 \\ \alpha_4^c & 0 \\ \alpha_5^c & 0 \\ \alpha_6^c & 0 \\ \alpha_7^c & 1 \\ \alpha_8^c & \alpha_8^{nc} \\ \alpha_9^c & \alpha_9^{nc} \end{pmatrix} \begin{matrix} \leftarrow \text{word analogy} \\ \leftarrow \text{arithmetic} \\ \leftarrow \text{matrices} \\ \leftarrow \text{grade German} \\ \leftarrow \text{grade foreign language} \\ \leftarrow \text{grade maths} \\ \leftarrow \text{conscientiousness} \\ \leftarrow \text{locus of control} \\ \leftarrow \text{self-esteem} \end{matrix} \tag{15}$$

These restrictions leave no rotation free. Let $\mathbf{A}^* = \mathbf{A}\mathbf{R}^{-1}$ and $\mathbf{\Sigma}_\theta^* = \mathbf{R}\mathbf{\Sigma}_\theta\mathbf{R}'$ be an observationally equivalent structure satisfying the same restrictions, and write $\mathbf{R}^{-1} = (r_{fg})$. The dedicated cognitive measurements require $\alpha_m^{*nc} = \alpha_m^c r_{12} = 0$ for $m \leq 6$, so $r_{12} = 0$. The two scale normalisations then give $r_{11} = 1$ and $r_{22} = 1$. Hence

$$\mathbf{R}^{-1} = \begin{pmatrix} 1 & 0 \\ r_{21} & 1 \end{pmatrix}, \qquad \mathbf{\Sigma}_\theta^* = \begin{pmatrix} \sigma_{\theta^c}^2 & -r_{21}\sigma_{\theta^c}^2 \\ -r_{21}\sigma_{\theta^c}^2 & \sigma_{\theta^{nc}}^2 + r_{21}^2\sigma_{\theta^c}^2 \end{pmatrix}.$$

The remaining transformation adds a multiple of the cognitive factor to the non-cognitive one, $\theta^{*nc} = \theta^{nc} - r_{21}\theta^c$, and shifts the cross-loadings of the non-cognitive measurements to $\alpha_m^{*c} = \alpha_m^c + r_{21}\alpha_m^{nc}$. Independence of the factors forces $r_{21} = 0$, so $\mathbf{R} = \mathbf{I}$. The alternative would be to restrict one non-cognitive measurement to be dedicated and let the factors correlate.

Order condition. The 36 off-diagonal covariances identify 12 parameters (ten free loadings and two factor variances), and the nine variances identify the nine uniquenesses. The system is therefore over-identified, with 24 testable restrictions.

Outcome loadings. Ignoring selection for intuition, the wage loadings would be recovered from

$$\mathrm{Cov}(\tilde{w}_i, \tilde{M}_1) = \alpha_w^c \sigma^2_{\theta^c}, \qquad \mathrm{Cov}(\tilde{w}_i, \tilde{M}_7) = \alpha_w^c \alpha_7^c \sigma^2_{\theta^c} + \alpha_w^{nc} \sigma^2_{\theta^{nc}},$$

where $\tilde{w}_i$ is the log wage net of observables, and similarly for the education and occupation equations. In the full model, the selection of individuals into education levels and occupations is accounted for by the likelihood (11).

C Estimation

C.1 First-stage likelihood

Define Θ the support of the latent factors $\theta = (\theta^{\mathbf{c}}, \theta^{\mathbf{nc}})$, and $\mathcal{M}_i$ the set of measurements observed for individual i. Integrating over the factor distribution, the likelihood for measurements writes

$$\mathcal{L}^M = \prod_i \int_\Theta \left[\prod_{m \in \mathcal{M}_i} f(M_{i,m} | \mathbf{S_i}, \mathbf{X_i}, \theta = \bar{\theta}; \beta_{\mathbf{m}}, \alpha_{\mathbf{m}}, \sigma_{\mathbf{M_m}}) \right] f_\theta(\bar{\theta}) d\bar{\theta}.$$

Given the normality assumption of the measurement idiosyncratic shocks,

$$f(M_{i,m} | \mathbf{S_i}, \mathbf{X_i}, \theta = \bar{\theta}) = \frac{1}{\sigma_{M_m}} \phi \left(\frac{M_{i,m} - \mu_i^m(\bar{\theta})}{\sigma_{M_m}} \right),$$
$$\mu_i^m(\bar{\theta}) = \beta_{0,m} + \mathbf{S_i}' \beta_{\mathbf{1,m}} + \mathbf{X_i}' \beta_{\mathbf{2,m}} + \alpha_{\mathbf{m}}^{\mathbf{c}} \bar{\theta}^{\mathbf{c}} + \alpha_{\mathbf{m}}^{\mathbf{nc}} \bar{\theta}^{\mathbf{nc}},$$

where σ_{M_m} is the standard deviation of the measurement shock and ϕ is the standard normal density function. Therefore, the first-step likelihood function becomes

$$\mathcal{L}^1 = \prod_i \int_\Theta \left[\prod_{m \in \mathcal{M}_i} \frac{1}{\sigma_{M_m}} \phi \left(\frac{M_{i,m} - (\beta_{0,m} + \mathbf{S_i}'\beta_{\mathbf{1,m}} + \mathbf{X_i}'\beta_{\mathbf{2,m}} + \alpha_{\mathbf{m}}^{\mathbf{c}}\bar{\theta}^{\mathbf{c}} + \alpha_{\mathbf{m}}^{\mathbf{nc}}\bar{\theta}^{\mathbf{nc}})}{\sigma_{M_m}} \right) \right] f_\theta(\bar{\theta}) d\bar{\theta},$$

which is Equation 7 in the main text.

C.2 Second-stage likelihood

The Gumbel assumption on the schooling utility shocks yields the multinomial logit contribution to the likelihood

$$f(D_i | \mathbf{S_i}, \mathbf{X_i}, \theta = \bar{\theta}; \beta_{\mathbf{D}}, \alpha_{\mathbf{D}}) = \prod_{j \in \mathcal{J}(\mathbf{S}_i)} \left[\frac{\exp(\tilde{U}_i^j)}{\sum_{k \in \mathcal{J}(\mathbf{S}_i)} \exp(\tilde{U}_i^k)} \right]^{\mathbb{1}\{D_i = j\}},$$

where $\beta_{\mathbf{D}}$ collects the intercepts and the coefficients on $\mathbf{S_i}$ and $\mathbf{X_i}$ in Equation 1, and $\alpha_{\mathbf{D}}$ the factor loadings $(\alpha^{c,j}, \alpha^{nc,j})_j$.

For the occupational choice, the Gumbel assumption yields the conditional logit probability that individual i chooses the observed occupation o_i, denoting $\tilde{U}_{io} = U_{io} - \varepsilon_{io}$,

$$f(o_i | \mathbf{P_i}, \mathbf{X_i}, \theta = \bar{\theta}; \delta, \mathbf{\Gamma}, \phi, \psi) = \frac{\exp(\tilde{U}_{io_i})}{\sum_{o' \in \mathcal{O}} \exp(\tilde{U}_{io'})},$$

where δ are the base task valuations, $\mathbf{\Gamma}$ the 3×2 matrix of factor loadings, and ϕ and ψ the coefficients on education and background interacted with tasks in Equation 2.

For the contribution of the wage equation, given the wage shock normality assumption, we

can write

$$f(w_i|\mathbf{P_i}, \mathbf{X_i}, \mathbf{o_i}, \theta = \bar{\theta}; \pi, \rho, \alpha_{\mathbf{w}}, \sigma_{\mathbf{w}}) = \frac{1}{\sigma_w}\phi\left(\frac{w_i - \mu_i^w(\bar{\theta})}{\sigma_w}\right),$$

$$\mu_i^w(\bar{\theta}) = \pi_0 + \mathbf{P_i}'\boldsymbol{\pi}_{\mathbf{P}} + \mathbf{X_i}'\boldsymbol{\pi}_{\mathbf{X}} + \rho^{\mathbf{c}}\mathbf{T}^{\mathbf{c}}_{\mathbf{o_i}} + \rho^{\mathbf{ip}}\mathbf{T}^{\mathbf{ip}}_{\mathbf{o_i}} + \rho^{\mathbf{m}}\mathbf{T}^{\mathbf{m}}_{\mathbf{o_i}}$$
$$+ \alpha_w^c\bar{\theta}^c + \alpha_w^{nc}\bar{\theta}^{nc},$$

where w_i is the log wage in occupation o_i and σ_w^2 is the variance of the wage shock. Similarly, denote $\hat{\mu}_i^m(\bar{\theta}) = \hat{\beta}_{0,m} + \mathbf{S_i}'\hat{\beta}_{\mathbf{1,m}} + \mathbf{X_i}'\hat{\beta}_{\mathbf{2,m}} + \hat{\alpha}^{\mathbf{c}}_{\mathbf{m}}\bar{\theta}^{\mathbf{c}} + \hat{\alpha}^{\mathbf{nc}}_{\mathbf{m}}\bar{\theta}^{\mathbf{nc}}$ the deterministic part of measurement m evaluated at the first-stage estimates.

Altogether, the second step becomes

$$\begin{aligned}
\mathcal{L}^2 = &\prod_i \int_\Theta \left[\prod_{m\in\mathcal{M}_i} f(M_{i,m}|\mathbf{S_i}, \mathbf{X_i}, \theta = \bar{\theta}; \hat{\beta}_{\mathbf{m}}, \hat{\alpha}_{\mathbf{m}}, \hat{\sigma}_{\mathbf{M_m}})\right] \\
&\times f(D_i|\mathbf{S_i}, \mathbf{X_i}, \theta = \bar{\theta}; \beta_{\mathbf{D}}, \alpha_{\mathbf{D}}) \\
&\times f(o_i|\mathbf{P_i}, \mathbf{X_i}, \theta = \bar{\theta}; \delta, \mathbf{\Gamma}, \phi, \psi) \\
&\times \left[f(w_i|\mathbf{P_i}, \mathbf{X_i}, \mathbf{o_i}, \theta = \bar{\theta}; \pi, \rho, \alpha_{\mathbf{w}}, \sigma_{\mathbf{w}})\right]^{\mathbb{1}\{w_i \text{ observed}\}} f_\theta(\bar{\theta})d\bar{\theta} \\
= &\prod_i \int_\Theta \left[\prod_{m\in\mathcal{M}_i} \frac{1}{\hat{\sigma}_{M_m}}\phi\left(\frac{M_{i,m} - \hat{\mu}_i^m(\bar{\theta})}{\hat{\sigma}_{M_m}}\right)\right] \\
&\times \prod_{j\in\mathcal{J}(\mathbf{S}_i)} \left[\frac{\exp\left(\tilde{U}_i^j\right)}{\sum_{k\in\mathcal{J}(\mathbf{S}_i)}\exp\left(\tilde{U}_i^k\right)}\right]^{\mathbb{1}\{D_i=j\}} \\
&\times \frac{\exp\left(\tilde{U}_{io_i}\right)}{\sum_{o'\in\mathcal{O}}\exp\left(\tilde{U}_{io'}\right)} \\
&\times \left[\frac{1}{\sigma_w}\phi\left(\frac{w_i - \mu_i^w(\bar{\theta})}{\sigma_w}\right)\right]^{\mathbb{1}\{w_i \text{ observed}\}} f_\theta(\bar{\theta})d\bar{\theta},
\end{aligned}$$

which is Equation 11 in the main text.

C.3 Gauss-Hermite quadrature

Gauss-Hermite quadrature provides weights ω_n and abscissæ ξ_n, $n = 1, \ldots, G$, for integral approximations of the form

$$\int_{-\infty}^{\infty} \exp(-\xi^2) g(\xi) d\xi \approx \sum_{n=1}^{G} \omega_n g(\xi_n),$$

where $g(\cdot)$ is any function weighted by $\exp(-\xi^2)$. The approximation is exact when g is a polynomial of degree at most $2G - 1$. The normality of the latent factors makes this rule particularly convenient, given the exponential term in the normal density. For $x \sim \mathcal{N}(\mu, \sigma^2)$, the change of variables $\xi = \frac{x-\mu}{\sqrt{2}\sigma}$, so that $dx = \sqrt{2}\sigma d\xi$, gives

$$\begin{aligned}\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) h(x) dx &= \frac{1}{\sqrt{\pi}} \int_{-\infty}^{\infty} \exp(-\xi^2) h(\mu + \sqrt{2}\sigma\xi) d\xi \\ &\approx \sum_{n=1}^{G} \frac{\omega_n}{\sqrt{\pi}} h(\mu + \sqrt{2}\sigma\xi_n).\end{aligned}$$

Since the mean of the factor distribution is not identified, we set $\mathbb{E}(\theta) = \mathbf{0}$. The scale of each factor is fixed by the loading normalisations $\alpha_1^c = 1$ and $\alpha_7^{nc} = 1$ (see Section 5.1), so the factor standard deviations σ_c and σ_{nc} are estimated. With independent normal factors, $f_\theta(\bar{\theta}) = \frac{1}{\sigma_c}\phi\left(\frac{\bar{\theta}^c}{\sigma_c}\right)\frac{1}{\sigma_{nc}}\phi\left(\frac{\bar{\theta}^{nc}}{\sigma_{nc}}\right)$, and applying the rule in each dimension gives

$$\int_{\Theta} h_i(\bar{\theta}) f_\theta(\bar{\theta}) d\bar{\theta} \approx \sum_{q_1=1}^{G} \sum_{q_2=1}^{G} \frac{\omega_{q_1}\omega_{q_2}}{\pi} h_i\left(\sqrt{2}\sigma_c\xi_{q_1}, \sqrt{2}\sigma_{nc}\xi_{q_2}\right),$$

a product grid of $Q = G^2$ nodes $\bar{\theta}_q$ with weights $\tilde{\omega}_q = \omega_{q_1}\omega_{q_2}/\pi$, as in Equation 12.

So we can rewrite and estimate the first stage as

$$\mathcal{L}^1 \approx \prod_i \sum_{q=1}^{Q} \tilde{\omega}_q \prod_{m \in \mathcal{M}_i} \frac{1}{\sigma_{M_m}} \phi\left(\frac{M_{i,m} - \mu_i^m(\bar{\theta}_q)}{\sigma_{M_m}}\right),$$

$$\mu_i^m(\bar{\theta}_q) = \beta_{0,m} + \mathbf{S_i}'\beta_{\mathbf{1,m}} + \mathbf{X_i}'\beta_{\mathbf{2,m}} + \alpha_{\mathbf{m}}^{\mathbf{c}}\sqrt{\mathbf{2}}\sigma_{\mathbf{c}}\xi_{\mathbf{q_1}} + \alpha_{\mathbf{m}}^{\mathbf{nc}}\sqrt{\mathbf{2}}\sigma_{\mathbf{nc}}\xi_{\mathbf{q_2}},$$

where σ_c and σ_{nc} are estimated jointly with the measurement parameters.

In the second stage, the nodes are evaluated at the first-stage estimates $\hat{\sigma}_c$ and $\hat{\sigma}_{nc}$, $\hat{\theta}_q = (\sqrt{2}\hat{\sigma}_c\xi_{q_1}, \sqrt{2}\hat{\sigma}_{nc}\xi_{q_2})$, and

$$\mathcal{L}^2 \approx \prod_i \sum_{q=1}^{Q} \underbrace{\tilde{\omega}_q \left[\prod_{m \in \mathcal{M}_i} \frac{1}{\hat{\sigma}_{M_m}} \phi\left(\frac{M_{i,m} - \hat{\mu}_i^m(\hat{\theta}_q)}{\hat{\sigma}_{M_m}}\right)\right]}_{\equiv \lambda_{iq}} \mathcal{P}_i^{D_i}(\hat{\theta}_q)\, \mathcal{P}_{io_i}(\hat{\theta}_q) \left[f(w_i|\hat{\theta}_q)\right]^{\mathbb{1}\{w_i \text{ observed}\}}.$$

The term λ_{iq} does not depend on the second-stage parameters. It is computed once from the first-stage estimates and acts as an individual-specific (unnormalised) posterior weight on node q.

There is no consensus about the number of nodes used to approximate the normal distribution: Ashworth et al. (2021) use seven, while Heckman, Humphries, et al. (2018) use 17. In practice, results are almost unaltered by this choice. We use $G = 20$ nodes per dimension, i.e. $Q = 400$ nodes in total.

D Additional Results

D.1 Measurement System

Table D1: Measurement equations

	Const.	Female	HH inc.	Rural	Realschule	Gymnasium	α_m^c	α_m^{nc}	σ_m	Signal
IST 2000R cognitive test										
Word analogy	−0.864***	−0.175***	0.015	−0.041	0.534***	1.244***	1	0	0.795	0.17
	(0.078)	(0.050)	(0.012)	(0.053)	(0.077)	(0.074)			(0.021)	
Arithmetic	−0.594***	−0.223***	−0.003	0.079	0.434***	0.877***	0.900***	0	0.914	0.11
	(0.087)	(0.056)	(0.013)	(0.059)	(0.085)	(0.083)	(0.115)		(0.022)	
Matrices	−0.853***	0.091*	0.028**	−0.043	0.551***	0.947***	0.761***	0	0.908	0.09
	(0.085)	(0.054)	(0.013)	(0.058)	(0.083)	(0.081)	(0.108)		(0.021)	
School grades										
Grade German	−0.377***	0.364***	−0.007	−0.066	0.126	0.430***	1.622***	0	0.770	0.37
	(0.091)	(0.059)	(0.015)	(0.063)	(0.087)	(0.087)	(0.213)		(0.028)	
Grade foreign lang.	−0.374***	0.293***	0.028*	−0.075	0.019	0.388***	1.519***	0	0.799	0.33
	(0.094)	(0.060)	(0.015)	(0.064)	(0.090)	(0.089)	(0.205)		(0.026)	
Grade maths	−0.252***	−0.089	0.027*	0.029	0.017	0.348***	1.617***	0	0.769	0.37
	(0.091)	(0.059)	(0.015)	(0.063)	(0.087)	(0.087)	(0.184)		(0.024)	
Non-cognitive measures										
Conscientiousness	−0.006	0.263***	−0.009	0.008	−0.085	−0.142	0.516***	1	0.904	0.17
	(0.100)	(0.063)	(0.015)	(0.068)	(0.098)	(0.095)	(0.133)		(0.028)	
Locus of control	−0.332***	−0.082	0.014	0.033	0.331***	0.359***	0.394***	1.169***	0.896	0.20
	(0.101)	(0.063)	(0.015)	(0.069)	(0.099)	(0.095)	(0.122)	(0.237)	(0.032)	
Self-esteem	−0.039	−0.323***	0.022	0.086	−0.053	0.184	0.305**	1.822***	0.720	0.46
	(0.124)	(0.078)	(0.018)	(0.085)	(0.123)	(0.117)	(0.146)	(0.424)	(0.078)	
Factor SDs: $\sigma_{\theta^c} = 0.365$ (0.036), $\sigma_{\theta^{nc}} = 0.361$ (0.054). $N = 1145$; log-likelihood −11697.69; $G = 20$ Gauss–Hermite nodes per factor.										

Notes: All nine indicators are standardised. Grades are recoded so that higher values are better grades. Reference school track is *Hauptschule*; household income is monthly, in 1,000€. The cognitive (non-cognitive) loading of word analogy (conscientiousness) is normalised to one. Missing indicators are integrated out of the likelihood (FIML). Signal is the share of the indicator's variance net of covariates that is explained by the two factors, $(\alpha_m^{c\,2}\sigma_{\theta^c}^2 + \alpha_m^{nc\,2}\sigma_{\theta^{nc}}^2)/(\alpha_m^{c\,2}\sigma_{\theta^c}^2 + \alpha_m^{nc\,2}\sigma_{\theta^{nc}}^2 + \sigma_m^2)$. Standard errors from the inverse Hessian; those of the standard deviations by the delta method. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

D.2 Decomposition of Occupation Wage Gaps

Each component is computed at the full-model estimates, on the workers with an observed wage. For worker i in occupation o, the task return is $\boldsymbol{\rho}'\mathbf{T}_o$, the education term $\mathbf{P}_i'\boldsymbol{\pi}_P$, the skill term $\alpha_w^c\,\mathbb{E}[\theta_i^c] + \alpha_w^{nc}\,\mathbb{E}[\theta_i^{nc}]$, with the expectations taken over the individual's first-stage posterior, and the background term $\mathbf{X}_i'\boldsymbol{\pi}_X$. Each term is averaged within the occupation and expressed as a deviation from its sample mean, as is the observed mean log wage (the wage gap). The residual is the wage gap minus the four components, i.e. the occupation's mean

wage residual. The variance shares quoted in the text are employment-weighted covariance shares, $\sum_o n_o\, c_o\, g_o / \sum_o n_o\, g_o^2$, where g_o is the wage gap, c_o a component and n_o the number of workers, so that the five shares sum to one.

Code	Occupation	N	Wage gap	Task return	Education	Skills	Background	Residual
27	Technical R&D	20	0.211	0.169	0.085	0.009	0.012	-0.063
26	Mechatronics & electr.	45	0.008	0.133	-0.016	0.003	-0.011	-0.101
40	Sciences & IT	40	0.196	0.132	0.060	0.005	0.005	-0.006
31	Architecture	9	0.362	0.098	0.148	0.006	0.050	0.060
72	Financial services	36	0.308	0.094	0.052	0.004	0.033	0.124
73	Law & public admin.	38	0.099	0.081	0.056	0.003	0.043	-0.083
25	Mechanical eng.	78	0.106	0.074	-0.048	-0.000	-0.020	0.099
23	Paper & printing	13	-0.039	0.051	-0.003	-0.001	-0.020	-0.065
53	Security	25	-0.107	0.046	-0.001	0.002	0.036	-0.191
52	Drivers	9	-0.236	0.038	-0.098	-0.010	-0.051	-0.115
90	Humanities & media	39	0.038	0.035	0.084	0.003	0.030	-0.114
71	Business mgmt.	80	0.199	0.032	0.054	0.004	0.033	0.076
51	Transport & logistics	53	-0.113	0.023	-0.088	0.003	-0.039	-0.012
24	Metal production	39	-0.126	0.017	-0.032	-0.004	-0.037	-0.070
61	Purchasing & trade	39	0.000	0.010	0.003	-0.001	0.033	-0.046
22	Plastics	20	-0.039	0.009	-0.061	-0.006	-0.066	0.086
34	Building services	19	0.069	0.005	-0.029	0.003	-0.010	0.101
81	Medical health	106	0.045	-0.012	0.020	-0.003	0.020	0.019
11	Agriculture	11	-0.157	-0.030	-0.008	0.005	-0.049	-0.075
54	Cleaning	5	-0.304	-0.030	-0.183	-0.013	-0.046	-0.032
28	Textiles & leather	1	-0.129	-0.039	-0.079	-0.021	0.001	0.010
29	Food production	36	-0.212	-0.045	-0.089	-0.001	-0.080	0.002
82	Non-medical health	39	-0.108	-0.050	-0.037	0.003	-0.033	0.010
84	Teaching	48	0.288	-0.055	0.151	0.006	0.061	0.124
32	Civil engineering	10	-0.042	-0.073	-0.077	-0.012	-0.048	0.168
83	Social work & household	74	0.055	-0.080	0.045	-0.007	0.017	0.079
1	Armed forces	13	-0.146	-0.084	-0.118	-0.001	-0.036	0.092
33	Interior construction	11	-0.017	-0.101	-0.011	-0.010	-0.069	0.174
63	Tourism & catering	57	-0.295	-0.105	-0.034	-0.002	0.002	-0.156
62	Retail sales	92	-0.282	-0.130	-0.064	-0.002	-0.014	-0.072
12	Gardening	9	-0.304	-0.182	-0.099	-0.009	-0.052	0.037

Table D2: Occupation mean log-wage gap (vs sample mean) decomposed into the occupation's own task return $\rho' T_o$ (education channel shut), education composition $E[P'\pi_P \mid o]$, skill sorting $E[\alpha_w^c \theta^c + \alpha_w^{nc} \theta^{nc} \mid o]$, background $E[X'\pi_X \mid o]$ and residual. Sorted by task return.

D.3 Wage Equation with Each Channel Removed

	Full	$\rho = 0$	$\pi_P = 0$
Intercept	2.110*** (0.062)	2.066*** (0.062)	2.210*** (0.040)
Education (base = Basic ed.)			
Gymnasium	-0.202** (0.093)	-0.203** (0.094)	—
Apprenticeship	0.105* (0.063)	0.131** (0.064)	—
Higher voc.	0.249*** (0.070)	0.282*** (0.071)	—
University	0.375*** (0.077)	0.401*** (0.078)	—
Background			
Female	0.020 (0.029)	-0.020 (0.027)	0.048 (0.030)
HH income (std.)	0.032** (0.014)	0.036** (0.014)	0.036** (0.014)
Rural	-0.030 (0.028)	-0.024 (0.028)	-0.022 (0.029)
Realschule	0.076* (0.043)	0.091** (0.043)	0.112*** (0.043)
Gymnasium	0.156*** (0.052)	0.185*** (0.051)	0.274*** (0.044)
Task returns			
Cognitive	0.136*** (0.039)	—	0.168*** (0.040)
Interpersonal	-0.110*** (0.025)	—	-0.116*** (0.026)
Manual	0.030 (0.037)	—	0.034 (0.038)
Factor loadings			
Cognitive factor	0.052 (0.052)	0.086* (0.052)	0.168*** (0.051)
Non-cognitive factor	0.066 (0.064)	0.111* (0.067)	0.045 (0.076)
Residual std. dev.	0.451	0.456	0.469
Log-likelihood	-17077.6	-17092.3	-17123.8

Table D3: Wage equation under channel restrictions, each re-estimated jointly with the education MNL and occupation clogit. Column $\rho = 0$ drops the occupation task returns: π_P is the total effect of education holding θ fixed. Column $\pi_P = 0$ drops the education dummies. One-step SEs from numerical Hessian (Stage-1 parameters treated as known). * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. $N = 1137$.