INSTITUTE OF ECONOMICS
Centre for Economic and Regional Studies

HUN-REN
Hungarian Research Network

Közgazdaság- és Regionális Tudományi Kutatóközpont
KÖZGAZDASÁG-TUDOMÁNYI INTÉZET

KRTK-KTI WORKING PAPERS | KRTK-KTI MŰHELYTANULMÁNYOK

Contribution of High School Heterogeneity to the Wage Variation of Young Workers

ISTVÁN BOZA – DÁNIEL HORN

KRTK-KTI WP – 2023/30
November 2023

INSTITUTE OF ECONOMICS, CENTRE FOR ECONOMIC AND REGIONAL STUDIES
HUNGARIAN RESEARCH NETWORK (HUN-REN)
BUDAPEST

ABSTRACT
The aim of this paper is to quantify how much of the initial wage differentials of young workers is explained by the secondary school they attended, and to disentangle the (descriptive) channels contributing to these differences. The analysis is based on the HUN-REN CERS Admin3 database, taking advantage of the fact that for some cohorts, young people's secondary schooling (and students' school standardized mathematics test scores) and wage outcomes at their early career can be observed simultaneously. Using wage decomposition methods, we separate the channels of firm and occupational selection from the direct returns to further education. Our analysis suggests that about 10 percent of the total wage dispersion of young people aged 18-25 (and already working) is generated at the school level. This also implies that the correlation between the wages of any two students of the same school is 0.1. Another novelty of the paper is that we show that a substantial part of these correlations are due to occupational and workplace selection (e.g. students from a given school type are systematically more likely to go on to well-paid jobs). If we remove these selection effects, the effect of schools on wage dispersion, the correlation between the latent skills of students, shrinks to 4 percent. Finally, we also compare schools of different quality based on different school characteristics (e.g. average test scores), which allows us to further stress the importance of the selection channels.

JEL codes: I24, I26, J31
Keywords: secondary schools; wage prospects; school quality; wage variation

István Boza
IE-CERS HUN-REN
boza.istvan@krtk.hun-ren.hu

Dániel Horn
IE-CERS HUN-REN
and Corvinus University of Budapest
horn.daniel@krtk.hun-ren.hu

Az iskolák heterogenitásának hozzájárulása a későbbi béregyenlőtlenségekhez a fiatal munkavállalók esetében

BOZA ISTVÁN – HORN DÁNIEL

ÖSSZEFOGLALÓ

E tanulmány célja, hogy számszerűsítse, hogy a fiatal munkavállalók kezdeti bérkülönbségének mekkora részét magyarázza az általuk látogatott középiskola, valamint, hogy szétválassza az e különbségekhez hozzájáruló csatornákat. Az elemzés a KRTK Admin3 adatbázisán alapul, kihasználva azt, hogy egyes kohorszok esetében a fiatalok középfokú iskolái (és a diákok iskolai standardizált matematikai teszteredményei) valamint a fiatalkori bérek egyidejűleg figyelhetők meg. Bérdekompozíciós módszerekkel szétválasztjuk a vállalati és a foglalkozási szelekció csatornáit a továbbtanulás közvetlen hozamától. Elemzésünk szerint a 18-25 éves (és már dolgozó) fiatalok teljes bérszóródásának mintegy 10 százaléka az iskolai szinten keletkezik. Ez azt is jelenti, hogy az ugyanazon iskola két tanulójának bére közötti korreláció 0,1. A tanulmány másik újdonsága, hogy megmutatjuk, hogy e korrelációk jelentős része a foglalkozási és munkahelyi szelekciónak köszönhető (pl. egy adott iskolatípus diákjai szisztematikusan nagyobb valószínűséggel jutnak el jól fizető munkahelyekre). Ha ezeket a szelekciós hatásokat eltávolítjuk, akkor az iskoláknak a bérszórásra gyakorolt hatása, a tanulók látens képességei közötti korreláció 4 százalékra zsugorodik. Végezetül, a különböző iskolai jellemzők (pl. átlagos teszteredmények) alapján is összehasonlítjuk a különböző minőségű iskolákat, ami lehetővé teszi számunkra, hogy még inkább hangsúlyozzuk a szelekciós csatornák fontosságát.

JEL: I24, I26, J31

Kulcsszavak: középiskolák, bérkilátások, iskolaminőség, bérszóródás

Contribution of High School Heterogeneity to the Wage Variation of Young Workers

István Boza[1] , Dániel Horn[1,2]

[1] Centre for Economic and Regional Studies, Budapest, Hungary
[2] Corvinus University of Budapest, Budapest, Hungary

November 8, 2023

The aim of this paper is to quantify how much of the initial wage differentials of young workers is explained by the secondary school they attended, and to disentangle the (descriptive) channels contributing to these differences. The analysis is based on the HUN-REN CERS Admin3 database, taking advantage of the fact that for some cohorts, young people's secondary schooling (and students' school standardized mathematics test scores) and wage outcomes at their early career can be observed simultaneously. Using wage decomposition methods, we separate the channels of firm and occupational selection from the direct returns to further education. Our analysis suggests that about 10 percent of the total wage dispersion of young people aged 18-25 (and already working) is generated at the school level. This also implies that the correlation between the wages of any two students of the same school is 0.1. Another novelty of the paper is that we show that a substantial part of these correlations are due to occupational and workplace selection (e.g. students from a given school type are systematically more likely to go on to well-paid jobs). If we remove these selection effects, the effect of schools on wage dispersion, the correlation between the latent skills of students, shrinks to 4 percent. Finally, we also compare schools of different quality based on different school characteristics (e.g. average test scores), which allows us to further stress the importance of the selection channels.

Keywords: secondary schools; wage prospects; school quality; wage variation
JEL codes: I24, I26, J31
Corresponding author: István Boza, istvan.boza@krtk.hun-ren.hu
The research has received funding from the National Research, Development and Innovation Office (grant no. NKIFH K-143415).

Introduction

Perhaps all social scientists agree that schools, whether elementary, secondary or tertiary, play a significant role in shaping the choices and behaviors of individuals as adults. This may affect both demographic (childbearing, health) or economic (labor market) outcomes. In addition to the acquisition of basic and vocational skills and competencies, schools also play a crucial role in the socialization of young people and the formation of social relationships. However, the "value added" of schools can be very different: their quality can vary, so that they can have a different impact on students' outcomes. This variation may depend on observed (e.g., provider of education, school type, region) and unobserved (e.g., teacher quality) factors. In addition to differences in school quality, the composition of students enrolled may also differ across schools, with these two factors being positively correlated, i.e. children with higher ability and family background are more likely to attend better schools. In fact, the coexistence of individual and school characteristics may be mutually reinforcing: for example, (better) schools may be able to create stronger networks of contacts, the effect of which may even be observed in the labor market.

It is therefore not surprising to expect that the future (labor market) position of students attending a given school is correlated: there are schools from which students systematically leave with better outcomes, and others from which they systematically leave with worse outcomes. Most studies in the literature focus primarily on outcomes related to further education. A summary of these outcomes can be found in Hermansen et al. (2020). Concerning future earnings, limited empirical evidence of school-level correlations is available, mostly from Scandinavian countries due to data requirements. Altonji and Mansfield (2011), using survey data, compared the adult wages of young individuals attending the same school earlier and found a correlation of 0.11 in the United States. That is the expected future wages of two randomly selected students from a given school correlate to a small, but non-negligible extent. In Sweden, this figure is 0.02 according to Lindahl (2011), while in Norway, it ranges from 0.02 to 0.06 as reported by Raaum et al. (2006) or even lower as found by Hermansen et al. (2020). After accounting for the effect of family background (i.e. school-level selection), all authors observe even weaker associations. These figures can also be interpreted as the extent to which schools contribute to wage inequality among young adults. Specifically, they also measure the degree to which the heterogeneity of school quality may explain the observed variation in wage levels of (young) adults.

In this paper, we first provide additional evidence on an explanatory role of secondary schools in future earnings variation, using high-quality linked employer-employee data from Hungary. Besides analyzing raw or partially adjusted wages, we also account for and analyze the channels of occupational and firm selection that may account for a significant portion of the returns to schooling (Cardoso et al., 2018). We quantify the contribution of these channels and the residual, individual-level wage component, that is the overall earnings potential (proxy of human capital) of students. We also assess the importance of these selection channels when comparing school of different type or quality (e.g. based on results on standardized test scores.) Our analysis suggests that about 10 percent of the total wage dispersion of young people aged 18-25 (and already working) is generated at the school level. If we remove the effects of firm and occupational selection, the contribution of secondary schools on wage dispersion shrinks to 4 percent. This implies important quality differences across schools, measured by this wage-oriented measure of school quality, capturing unobserved future labor market potential. By controlling for standardized test scores measured around the age of 16, we can get an even more nuanced measure of residual school quality, suggesting differences in schools' ability to provide valuable (non-cognitive) soft skills for their pupils.

The descriptive finding that students with good wage prospects bunch together into specific high schools could be the result of multiple (non-exclusive) mechanisms. First, different schools are expected to enhance student skills to a different extent, and the differences in school value added could be the main driver of differences in either educational outcomes – e.g. standardized test scores – or future labor market outcomes. Second, the non-random admission of students into secondary education could also lead to heterogeneity, even in itself. Some schools may simply be able to pool together individuals of high skills, without substantially enhancing their earnings potential. Third, high schools can also heterogeneously affect labor market outcomes through the different level of social capital their students accumulate. As recent studies show, former co-workers or university peers can both provide better labor market prospects for their acquaintances (Boza & Ilyés, 2020; Ilyés & Sebők, 2020). Even after accounting for the former differences, students may be clustered in schools based on their unobservable, non-cognitive or soft skills. Educational units could thus also provide signals for potential employers with respect to these, otherwise unobservable characteristics. Whether some (elite) high schools could exert a signaling effect on the labour market – on top of the higher education degrees their alumni tends to obtain – is a novel empirical question.

In the second part of the paper we aim to further our knowledge about the schooling system by focusing on the above-listed, potential channels in which high-schools can shape labor market outcomes. We propose a methodological framework that could capture not only the overall effect of schools on wage outcomes, but would also allow us to decompose these gains into channels attributable to heterogeneity in value added, selection of student pools, and the joint effects of networks and signaling. Due to data limitation, the empirical findings regarding this exercise are limited, and will need further analysis in the future.

The direct comparison of education data and future labor market outcomes is made possible in the Admin3 dataset, in which we can not only observe the full employment histories of individuals, but for more recent cohorts we have enough information on educational factors as well. Most notably we know which students had attended the same schools, and also we have the results of the National Assessment of Basic Competences (NABC) – a standardized test of mathematics and literacy skills – linked to the individuals. So far Hermann et al. (2022) presented results on the positive relation between such test scores and future wages and employment probabilities. Also, Boza (2021, 2022) showed that this relation between scores (both in absolute or within-school relative terms) and future wages is not solely driven by the better cognitive skills of students. The systemic sorting of high-achieving students into high-wage and high-productivity firms and high-wage occupations amplifies wage differences, revealing a prominent role of assortative matching in the labor market. In this paper we build on these findings, and the related methods (multi-way high dimensional fixed effects approach) as well.

Our contribution to the literature is twofold. First, we provide descriptive evidence on the direct and indirect contribution of secondary school heterogeneity to future wage prospects, using data from Central Eastern Europe. When doing so, compared to previous studies we differentiate between the sorting mechanisms, and those differences that relate to portable skills of the students. Second, after documenting a non-negligible variation in school quality, we aim to decompose these differences into factors relating to pre-admittance sorting and school value added with respect to observable (cognitive) skills and as a residual part, the contribution of (non-cognitive) soft skills.

The study is structured as follows. Section 1 presents that dataset used and the country-specific context. Section 2 presents the applied methods for the different exercises of the study. Section 3 discusses results and Section 4 concludes.

1 Data and Background

Schools in Hungary

The Hungarian schooling system consists of three tiers, and although some elements went through reforms during the past two decades, the basic distinctions remain the same. Elementary education generally spans eight years, from the age of 6 or 7 to the age of fourteen/fifteen. After 8^{th} grade – or in the case of some high schools 4^{th} or 6^{th} –, students may choose general high schools (”gimnázium”) or vocational schools. High schools and a part of vocational schools (”szakközépiskola”) have a graduation test in 12^{th} grade, called ”érettségi”. This is a requirement for entering higher education, which mostly consists of college and university programs of 3 bachelor and 2 master years according to the Bologna system. One cohort of students in secondary schools is about 100 000 students – of which population we can observe around 50 000 individuals per cohort.

Linked Employer-Employee Data

The empirical estimations in this study use data from the Databank of the HUN-REN Research Centre for Economic and Regional Studies. The Panel of Administrative Data ('Admin3') from CERS is a large, administrative, linked employer-employee panel dataset, covering a random fifty percent of the Hungarian population. The two-way panel spans from 2003 through 2017 and contains labor market data in monthly resolution, such as an ID for the employer, earnings in given month, occupation information[1] and balance sheet data for incorporated employers.[2] We observe all taxed earnings from the given employer during the given month, but cannot differentiate between bonuses, and general wage.[3] The data does not convey any family-related information, only individual characteristics like gender, age, residence and also some variables on healthcare expenditures and specific transfers received by the individuals, which we do not utilize in this research. Although we do not have a common ”highest education” variable available for the full panel, in the second half of the observation period, that is for the population of interest in this study, we have exact information on the high school and university attendance of the individuals.

Data on schools and performance

Besides the standard labor market information available in common linked employer-employee panels, Admin3 has a number of additional, often unique, modules of interest. For this study, we are utilizing the availability of results of the National Assessment of Basic Competencies (NABC), which allows for identifying students attending the same educational institutions. Results for this standardized test score, along with the associated (anonymized) school identifiers are available for certain cohorts of students starting with those who took the test in 2008. This allows us to analyze variations in wages for at least one full year spent in the labor market.

Our focus is on secondary schools, particularly those institutions where individuals had their most recent tenth-grade standardized examination. For simplicity, we do not differentiate between sites with different physical addresses within an institution. International studies usually analyze outcomes between the ages of 30-35, but unfortunately we do not (yet) have that possibility. The earliest data we have available is from the class of the 10^{th} grade in 2008, who are only 26-27 years old in 2017, the final year of our dataset. To ensure comparability, we evaluate four cohorts in total, including the 2009-2011 cohorts besides the 2008 one. These students could spend up to maximum seven years in the labor market – at least those members of the earliest cohorts who did not enter higher education. We analyze earnings outcomes from the ages of 18 to 25 and focus solely on wages from each October from 2010 onward. We have to note that our estimates of the contribution

[1]Occupational heterogeneity throughout our paper refers to 332 occupational categories obtained from the harmonization of two sets of 4-digit occupations based on the Hungarian equivalent of the ISCO system – one for the years before 2012 and one thereafter.

[2]Unlike LEED data from many other countries we lack establishment identifiers, so we can treat only whole firms and institutions as the unit of observations.

[3]The social contribution reports which form the basis of the data have to be submitted on a monthly basis since 2012. Before that, yearly earnings from an employer were attributed to calendar months accordingly to the number of days of the employment spell belonging to the given month.

of tertiary education are limited due to the fact that not all of our study participants had completed their education at the time of our analysis. [4] Therefore, our results are primarily driven by the labor market outcomes of individuals who either did not complete tertiary education or only spent a brief period in higher education.

2 Methods

2.1 Estimating wage components in a multi-stage model

For our main findings we rely on the methodological framework that builds upon the seminal work of Abowd et al. (1999). Specifically, we start by estimating the following log-additive model of wages, following Boza (2022) and Torres et al. (2018), on the full timespan of the Admin3 dataset.

$$\ln w_{ijt} = \boldsymbol{X}_{ijt}\boldsymbol{\beta} + \psi_j + \lambda_{k(ijt)} + \theta_i + \varepsilon_{ijt} \tag{1}$$

In this model, while ψ_j captures time-invariant firm premia – how the given firm over or underpays all of its workers compared to the market – and λ_k captures occupational heterogeneity, the individual-level θ_i parameters will capture the earnings potential of all individuals. By the logic of the AKM framework this parameter encompasses all the time-invariant individual traits that are valuated (in wage terms) in the labor market, regardless of which firms or occupations the individual would work in. Hence, these terms proxy both observed and unobserved, both cognitive and non-cognitive skills as well. X_{ijt} captures the effect of some time-variant and job-specific factors, such as age, tenure or calendar year, and also whether the individual is contracted in the private or the public sector. In a second stage, the wage effect of observed and unobserved firm-level and personal characteristics could be also disentangled.

$$\hat{\psi}_j = \boldsymbol{Z}_j\boldsymbol{\gamma} + \varepsilon_j^J \tag{2}$$

$$\hat{\theta}_i = \boldsymbol{W}_i\boldsymbol{\eta} + \varepsilon_i^I \tag{3}$$

Regarding firms effects, we can remove industry wage differentials and the wage effects of foreign or state ownership. The residual firm effect, ε_j^J will thus capture the within-industry, relative premia of the given firm. Similarly, from the individual effect (general worker quality), we can remove the effects of birth cohort, residence (modal region) and gender, so that ε_i^I could be used for comparing individuals of the same age, location and gender. Another such time-invariant trait could be the identity of the high-school the individual has attended or even their cognitive skills at a given point in time – which we aim to measure by the 10^{th} grade NABC scores. Before taking these test scores into account, we augment the corresponding second-stage regression to incorporate fixed effects for all the secondary schools present in our data , indexed by s. Thus, we will arrive at a novel estimator of school quality.

$$\hat{\theta}_i = \boldsymbol{W}_i\boldsymbol{\eta} + \xi_{s(i)} + \varepsilon_i^I \tag{4}$$

In this formulation, the ξ_s parameters will reflect the average earnings potential of students of the given school, s – as measured by the AKM person effects –, controlled for composition with respect to variables in $\boldsymbol{W}_i$, gender, region and cohort. Schools with a high ξ_s parameter either pool together or train pupils that can expect to earn relatively higher wages over others at any given potential employer.

[4] While we have test score data available for 6.85% of the individuals in our sample, this corresponds only to 1.67% of total wage observations.

The ability to observe standardized test scores in our data as a proxy of cognitive skills provides us with an opportunity to provide more detailed estimations of school effects and its sub-components. Thus, we augment Equation 4, by incorporating terms capturing the performance of individuals relative to their school average regarding their test scores and – optionally – their educational attainment. If we also include a variable for future higher education attainment in $\boldsymbol{W}_i$, we can also remove distinguish an important channel of school-related gains, further education. This may or may not be desired, depending on the type of analysis.

$$\hat{\theta}_i = \underbrace{\boldsymbol{W}_i\boldsymbol{\eta}}_{\text{Comp.}} + \underbrace{\beta_S(\text{Score10}_i - \overline{\text{Score10}}_{s(i)})}_{\text{Returns to (rel.) skills}} + \underbrace{\left[\beta_H(\text{H}_i - \overline{\text{H}}_{s(i)})\right]}_{\text{Returns to (rel.) educ.}} + \underbrace{\xi_{s(i)}}_{\text{Total school effect}} + \varepsilon_i^I \tag{5}$$

The school-level ξ_s parameters of this formulation will still account for composition with respect to some observational traits. Due to the fixed effect design, the β parameters are estimated from within-school variation[5], with β_S capturing that students better in their schools will have higher earnings potential at all employers, similar as those who study more in the future – captured by β_H. However, due to the formulation in relative terms, between school differences in average skills (or education) will be captured by the school effect parameters. Therefore these effects will still convey information on both the average level of returns to cognitive and non-cognitive skills for the students for the given school.[6] An alternative formulation, however, without demeaning would provide a school effect net of both skill composition and future outcomes.

$$\hat{\theta}_i = \underbrace{\boldsymbol{W}_i\boldsymbol{\eta}}_{\text{Comp.}} + \underbrace{\beta_S\text{Score10}_i}_{\text{Returns to skills}} + \underbrace{[\beta_H\text{H}_i]}_{\text{Returns to educ.}} + \underbrace{\xi^*_{s(i)}}_{\text{Adj. school effect}} + \varepsilon_i^I \tag{6}$$

Let us call this kind of school effect, ξ^*_s an adjusted one, reflecting the fact that the skill composition of schools (as a possible source of sorting) is removed from the given term. This parameter would thus reflect the average level of non-cognitive traits, or at least traits independent of the observed test scores, in the given school. Again the returns to higher education (what remains after taking out occupational and firm-level sorting) could be captured by including a set of educational (dummy) variables. Combining equations 1, 2 and 6, we get the following, detailed decomposition of any wage observation, which we will use throughout various exercises of this study.

$$\ln w_{ijt} = \boldsymbol{X}_{ijt}\boldsymbol{\beta} + \boldsymbol{Z}_j\boldsymbol{\gamma} + \varepsilon_j^J + \lambda_{k(ijt)} + \boldsymbol{W}_i\boldsymbol{\eta} + \beta_S\text{Score10}_i + \beta_H\text{H}_i + \xi^*_{s(i)} + \varepsilon_i^I + \varepsilon_{ijt} \tag{7}$$

2.2 Constructed measures of labor market performance

For our analysis, starting from Equation 7, we define some specific wage outcomes of interest. We start from the observed total (raw) log wage ($\ln w$), and then subtract the estimated impact of selected factors in order to identify the channels through which differences between schools may emerge. The factors may be extracted in varying orders, but we follow the following logic to ensure clarity and consistency.

First, we subtract the effects of basic demographic characteristics, such as gender and age, from the total wage. Next, we eliminate the impacts of tenure, calendar year, and worker location to account for temporal and regional variations. The resulting metric may be referred to as a basic conditional wage, as it enables more informative comparison than the raw wage in many instances, since it is independent of the individual's gender and

[5]This would be true, even if we would not demean the variables, due to the within-school estimator implied by the inclusion of school fixed effects.

[6]Still, the inclusion of the relative deviations is important as those factors could be correlated with other components of W_i such as gender and region.

age, local labor market characteristics (place of residence), and macroeconomic conditions (time). Using the above notation, this measure is defined as:

$$\underbrace{w^*_{ijt}}_{\text{Adj. wage}} = \underbrace{\ln w_{ijt}}_{\text{Log. wage}} - \underbrace{\boldsymbol{W}_i\boldsymbol{\eta}}_{\text{Age, gender}} - \underbrace{\boldsymbol{X}^I_{it}\boldsymbol{\beta}^I}_{\text{Tenure, region, year}} \tag{8}$$

Next, we extract the observable (sector, size) and unobserved but time-constant characteristics of firms (residual firm effects) sequentially, followed by the impact of occupations (and public/private sector) and temporal variations. The resulting, time-invariant wage component measures the earning potential of individuals, i.e. how much more they would be expected to earn relative to the average worker (of the same sex and age) in any given firm and occupation under the same conditions. Essentially, it is a basic measure of human capital in economic terms.[7] In our notation:

$$\underbrace{\theta^*_i}_{\text{Earnings potential}} = \underbrace{w^*_{ijt}}_{\text{Adj. wage}} - \underbrace{(\boldsymbol{Z}_j\boldsymbol{\gamma} + \boldsymbol{X}^J_{jt}\boldsymbol{\beta}^J + \varepsilon^J_j)}_{\text{Industry, size, ownership, unobs.}} - \underbrace{(\lambda_{k(ijt)} + \boldsymbol{X}^K_{k(ijt)}\boldsymbol{\beta}^K)}_{\text{Occupation, contract}} - \underbrace{\varepsilon_{ijt}}_{\text{Unexp. resid.}} \tag{9}$$

In the last step, the residual component of an individual's wage is decomposed into three factors. First, the earnings potential explained by educational attainment. Second, the average levels captured by schools from the parts not explained by education, or skill composition (adj. school fixed effects). And finally, individual deviations from the school measures (independent, random factors). In an extended version of our model, we attempt to distinguish the impact of variations in cognitive ability by using the math score results from the NABC test in tenth grade. Still, simply incorporating the tenth-grade scores does not allow for differentiating between secondary school selection and school value-added – that task is left for an even more complex model. However, in this model, we already attempt to comprehend the influence of education on non-cognitive abilities net of student's observable skill composition. That is we aim to assess the relevance of soft skills, including that of social capital accumulation and networking. Our most detailed decomposition could be thus written as:

$$\underbrace{\xi^*_{s(i)}}_{\text{[Adj.] school effect}} = \underbrace{\theta^*_i}_{\text{Earnings potential}} - \underbrace{w^*_{ijt}}_{\text{Adj. wage}} - \underbrace{[\beta_S \text{Score10}_i]}_{\text{Returns to math}} - \underbrace{[\beta_H \text{H}_i]}_{\text{Returns to educ.}} - \underbrace{\varepsilon^I_i}_{\text{Unexp. ind.}} . \tag{10}$$

Within this framework, in the first half of the paper, we will address three questions. First, we look descriptively at how much of the earnings variation of young workers in the sample are explained by the above detailed components. Second, we will determine the proportion of total wages, adjusted wages and individual earning potential dispersion explained by a simple model encompassing solely school effects. This exercise will be (almost) identical to the estimation of intraclass correlations. Finally, we will briefly evaluate the magnitude and channels through which differences between schools of different type or quality may arise.

2.2.1 Contribution of school related components to total wage variation

Provided that all the above wage components are properly estimated, we can use the following sample moments coming from the variance-covariance matrix of the estimated parameters to characterize the contribution of secondary schools to future wages:

[7] It is different from the person effect of the AKM model, as the effects of gender and birth cohorts are already removed from this component.

- $\frac{\text{cov}(\hat{\xi}^*_{s(i)}),\ln w_{ijt})}{\text{var}(\ln w_{ijt})}$: The total contribution of differences in school quality (as we measure it) to the total wage variation in our sample with wage data. This is a component from the ensemble decomposition of Card et al. (2018). Naturally the contribution of other wage components could be calculated the same way. This parameter captures how less would wages vary in a counterfactual state, in which all school would provide (or cluster) the same (soft) skills.

- $\frac{\text{var}(\hat{\xi}^*_{s(i)})}{\text{var}(\ln w_{ijt})}$: The direct contribution of school heterogeneity to future wage variation. This is complemented by sorting channels, such as:

- $\text{corr}(\hat{\xi}^*_{s(i)}, \hat{\psi}_j) \rightarrow \frac{2\text{cov}(\hat{\xi}^*_{s(i)},\hat{\psi}_j)}{\text{var}(\ln w_{ijt})}$: this parameter captures how smaller wage variation would be if there would not be a (positive) correlation between school quality and future firm quality.

- $\text{corr}(\hat{\xi}^*_{s(i)}, \hat{\lambda}_k) \rightarrow \frac{2\text{cov}(\hat{\xi}^*_{s(i)},\hat{\lambda}_k)}{\text{var}(\ln w_{ijt})}$: Finally, this parameter captures how smaller wage variation would be if there would not be any connection between school quality and occupational choice.

2.2.2 Role of high school heterogeneity

In our second exercise, we take a step back. In our previous, total decomposition, the direct contributions of schools that we may find would only reflect how students of similar skills or even of similar soft skills are clustered together in schools. However, we can expect that better schools do not only have a direct contribution through increasing human capital, but through the various selection channels as well.

To characterize the contribution of schools to wage outcomes, previous scholars in the literature, for instance Hermansen et al. (2020) relied on estimating intraclass correlation measures. Given a single random effects regression model, with an outcome (say wages) on the left hand side, and a full set of school identifiers on the right hand side, the correlation in the expected outcome of two students of the same school would be given by the *ICC* formula.

$$y_{is} = \beta_0 + s_s + e_{is} \tag{11}$$

$$ICC_s = \frac{\sigma_s^2}{\sigma_s^2 + \sigma_e^2} = \frac{\text{var}(\hat{s}_s)}{\text{var}(y_{is})} = \frac{\text{cov}(\hat{s}_s, y_{is})}{\text{var}(y_{is})} = \text{contr}_s \tag{12}$$

As we indicate in Equation 12, the variance explained by the estimated school effects is analogous with the *ICC* interpretation. That is, in a simple formulation the question whether how much school heterogeneity contributes to wage variation, and the question of to what extent the wages of two students of the same school correlate is the same. In this paper we will estimate this measure for all the wage measures we defined earlier, across Equations 8 and 10.

If we want to extend this approach to account for differences in the skill composition of different schools, we can take the following approach. While we keep various outcomes on the left hand side, on the right hand side besides the school effects we include additional regressors.

$$y_{is} = \beta_0 + s_s + \beta_T \times \text{TEST}_i + e_{is} \tag{13}$$

In this setup, the covariance of the estimated effects and the outcome will be still informative – showing the contribution of school heterogeneity to wage variation, conditional on the student pool. However, it will not be equal to the intraclass correlation, of two random students, as skill levels within the same schools could be correlated as well.[8] A solution for that is to focus on adjusted outcomes, that is remove the effect of score composition differences first, and then use the usual ICC formula the following way.

$$y^*_{is} = y_{is} - \beta_T \times \text{TEST}_i = \beta_0 + s_s + e_{is} \tag{14}$$

$$ICC^*_s = \frac{\text{cov}(\hat{s}_s, y^*_{is})}{\text{var}(y^*_{is})} \neq \frac{\text{cov}(\hat{s}_s, y_{is})}{\text{var}(y_{is})} = \text{contr}^*_s \tag{15}$$

2.2.3 Differences between school types

Building on the works of Cardoso et al. (2016), Cardoso et al. (2018) and Boza (2021) we can also assess the channels through which wage differentials between different school types are generated, through the lens of the (augmented) AKM model. Notably, the authors propose, that wage differences across an observable characteristic, such as gender, education or firm ownership, the associated wage components could be unambiguously decomposed into differences in whether the affected observations belong to higher wage individuals, firms or occupations. Due to the linear nature of the AKM model, the following decomposition will hold for any variable C included either in the control variables or nested within the time-invariant person, firm or occupation elements.

$$\frac{\partial \text{ln} w_{ijt}}{\partial C} - \frac{\partial \boldsymbol{X_{ijt}\beta}}{\partial C} = \frac{\partial(\theta_i - \xi_{s(i)})}{\partial C} + \frac{\partial \xi_{s(i)}}{\partial C} + \frac{\partial \psi_j}{\partial C} + \frac{\partial \lambda_{k(it)}}{\partial C} \tag{16}$$

As our second stage decomposition incorporating school effects as presented in Equation 5, we may allow for some of the individual-level differences to be generated either within or between high-schools, as in the above formula. We could also rely on Equation 7, and take the differences along C in all the different sub-components. These differences will also add up to the total difference in log-wages.

Depending on our choice of C, we may be able to observe various patterns governing labor market allocation and wage formation. For instance, we can choose C to be the NABC test score themselves – a relation that Boza (2021) investigates graphically – to show that higher skill individuals get into better occupations and also end up with better firms. Also we could see whether these students tend to earn higher wages everywhere, and whether they tend to bunch together in schools that provide above average wage prospects.

However, we could also classify schools based on their baseline student selection, school value-added or even the unexplained school effect components, and use this classification as running variable C. Therefore we may reveal, that schools that are considered good due to value added and schools that simply pool together high-achieving students may have different effects on occupational choice, or worker sorting to better firms. A positive relation between the unexplained school component and firm effects would provide some suggestive evidence for the role of social capital.

Finally, differences across regions, the operator/financier of given schools or the curriculum could be used to provide policy relevant implications, about whether, and how wage differentials across these dimensions are generated. For instance vocational schools may have a stronger effect on allocation between sectors or occupations, while general high schools funnel students into better firms within any sectors, or provide high expected wages everywhere. In the current working paper, we only present preliminary results related to some school-level quality measures.

[8] The correlation would be given by the formula: $ICC_s = \frac{\sigma_s^2 + 2\sigma_{sT} + \sigma_T^2}{\sigma_s^2 + 2\sigma_{sT} + \sigma_T^2 + \sigma_e^2}$.

2.3 Sources of high school heterogeneity

Our proposed model for decomposing the school effect into components relating to the different possible channels presented in the introduction takes the following form. [9] It is important to note that here we use the total school effect from Equation 5, that still includes the mean effect of student composition.

$$\xi_{s(i)} = + \underbrace{\beta_S(\text{Score6}_i)}_{\text{Selection}} + \underbrace{\beta_{VA}(\text{Score10}_i - \text{Score6}_i)}_{\text{Returns to value added}} + \\ \underbrace{[\beta_H \text{H}_i]}_{\text{Returns to HE}} + \underbrace{[\beta_T T_{s(i)}]}_{\text{School tpye}} + \underbrace{\varepsilon_{si}}_{\text{Signaling, etc.}} \quad (17)$$

Equation 17 builds on our assumption that high schools may have two major source of variation that shapes the future outcomes of their students. First, we allow schools to have different pools of students both regarding some observable background traits – which should be mostly eliminated in the previous stage – and notably in the cognitive skills of their pupils. We aim to capture this by accounting for high school skill composition based on earlier test results of the students. As individual test scores from different years probably highly correlate, β_S should capture most of the returns to skill composition of the given school.[10] Second, we would like to capture the returns to school-level value added, which we capture with a very crude measure, the four-year increase in standardized scores for current pupils of the schools.

Although for both of these channels we assume a homogeneous (return) parameter, schools are heterogeneous both with respect the initial skills of their students and in what extent these skill have developed during the first two years.[11] The unexplained variation that remains in the school affect for accounting for these mechanisms will capture all the future gains for students that do not relate to these cognitive skills. Accordingly, we assume that the remaining school-level residuals may reflect on one hand the above/below average endowment of student with soft skills at the given school, and on the other hand some raw signal power or the school or accumulated social capital.

The components of the above decomposition could be utilized in a covariance decomposition to provide insight about the relative importance of the different channels in the heterogeneity of the returns of different schools, and more generally to the heterogeneity in earnings potentials (the person effects) or total wages. We note, that for detailed assessment, we can also take into account the future education of students and also allow for baseline differences – or even heterogeneous parameters – across different school types, with respect to the education provider or their curriculum.[12]

[9] Although the model is formulated on the individual's level, an identical alternative specification could be written up using school level averages of variables and proper weighting. These estimation still rely on future wage observations of those who work, and for the decomposition identities to hold, individuals who work for longer periods have to enter the estimations with higher weights.

[10] This, however, will limit the set of cohorts we can apply our decompositions to, as currently only a few cohorts have primary school test scores and labour market outcomes at the same time. Using eight grade scores is an alternative option.

[11] Unfortunately there is no test in 12^{th} grade, so the more prominent part of secondary education could not be assessed with this measure.

[12] Also, the variables used here along with the obtained residuals could be used for assessing the composition of future wage gains not only trough the person effect channel, but through sorting into higher wage firms or better occupations. Given the current data limitation, such detailed models are out of the scope of the current manuscript.

3 Results

3.1 Contribution to wage variation

The columns in Table 1 quantify (in percentage form) the contribution of each component to total wage dispersion in our sample, based on three different specifications. While in the first column we do not account for future educational attainment nor skill composition of schools in the final step of our wage decomposition approach, we add these factors in subsequently through the second and third specifications.

It can be seen that wages at a young age are significantly affected both by the occupational choice, and especially strongly by the firm (industry) where the students end up. While demographic characteristics explain 16 percent of the variance in wages, firm and occupation related characteristics explain 68 percent. However, school-related characteristics and the individual's earnings potential explain only 15 percent of the variance in total. As the last row shows, only 1 percent of this variance is systematically observed at the school level. [13] Independent of whether we include test scores and education (and at the same time narrow the sample to those for whom a tenth-grade maths score is available), we arrive at a very similar decomposition.

Table 1: The contribution of additive wage components to total wage dispersion (percentage shares)

	Baseline model	With HE	With NABC scores
Log wage (Variance: 0.159)	100.0	100.0	100.0
Gender and age	4.5	4.5	4.4
Tenure, region, year	12.1	12.1	12.0
Adjusted wage			
Firm (ownership, industry, size, unobserved)	40.9	40.9	40.9
Occupation (and sector)	10.4	10.4	10.4
Other unexplained	17.0	17.0	17.1
Personal earnings potential			
Educational attainment	-	1.3	0.9
10th grade math score	-	–	1.2
Unexplained individual	13.4	12.5	12.3
High school (id)	1.9	1.4	0.9
Additional moments			
$2\text{Cov}(\xi^*_{s(i)},\psi_j)$	1.3	0.8	0.5
$2\text{Cov}(\xi^*_{s(i)},\lambda_k)$	0.7	0.4	0.0
$\text{Var}(\xi^*_{s(i)})$	0.7	0.6	0.5

The last rows (the bottom most panel) of the table highlight, that the direct effect of schools is complemented by important selection channels. Notably, we could say that if schools would not contribute to assortative matching (sending high potential workers to the best firms), then the total wage variation around the age of 25 would be 0.5-1% smaller.

3.2 Intraclass correlations

The columns in Table 2 present, in order our composite wage measures defined through Equations 8 and 10 and show how much of the variance of the given indicator can be explained by secondary school heterogeneity. This indicator also corresponds to the intraclass

[13]For comparison, based on Boza (2021), individual characteristics explain around 50 percent of the total wage dispersion in the total population, firm characteristics around 22 percent, and occupations around 8 percent.)

correlation (ICC) in a simple model. In Specification 1, we first focus on such a model, in which we do not remove the additional effect of test scores on the given wage measures. The 11.3 percent in the first row can be interpreted in two ways. On the one hand, we can say that school heterogeneity – i.e. the fact that secondary schools differ from each other in all sorts of observable and unobservable characteristics – can explain a little more than 11 percent of the total dispersion in early-career wages. This also means that comparing arbitrary students with random school mates would yield a correlation of 0.11 between the two students' 18-25 year old wage levels. This figure is roughly in line with the U.S. results of Altonji and Mansfield (2011), but higher than results from Nordic countries.

Table 2: The shares of each wage indicator explained by school heterogeneity (percentage, or intraclass correlation × 100)

	(1)	(2)	(3)
	Model with HE	With NABC scores	
	$ICC_s = \text{contr}_s$	ICC_s^*	contr_s^*
Log wage	11.3	7.1	8.3
Adjusted wage	11.3	7.5	8.6
Without firm selection	6.8	4.4	5.2
Without firm and occupational selection	2.6	1.9	2.1
Personal earnings potential	4.6	3.5	3.8
Without effect of further education	3.7	3.2	3.4
Without effect of math skills	-	3.2	3.3

In the subsequent rows, we present the wage dispersion explained by high school heterogeneity after adjusting for common wage components. In the second row, we obtain an indicator very similar to the one above, after filtering out the effect of general characteristics (gender, age at birth, region, year, length of employment). However, our results show that a significant share of the differences also arise in this context due to to firm-specific selection. For example, when we look at within-firm wage dispersion (row 3), schools explain only 6.8 percent of the future wage dispersion. This decrease is possible if school choice is correlated with subsequent firm choice: that is if there are schools from which students are systematically placed in better firms and schools from which students are systematically placed in worse firms.

The fact that students from different educational institutions consistently choose different occupations (with different wages) also seems to be an important source of wage differences. For a particular occupation (and with similar average firm wages), only 2.6 percent of the wages of young workers can be explained by who studied in which secondary school. This is not unexpected, given that vocational and upper secondary schools are both included in the sample. Of the total dispersion of individual earnings potential (individual fixed effects, which embody human capital and are not explained by other time-varying factors), slightly more than 4.6 percent can be attributed to differences in secondary schooling.[14] Around one-fifth of this dispersion can be eliminated by including the highest education attained by students (3.7 percent instead of 4.6 percent). By filtering out the channel of higher education, we can say that schools can still explain 3.7 percent of the wage dispersion independent from education and subsequent selection channels. In other words, we also observe a correlation of around 0.04 between the raw earnings potential of two arbitrary schoolmates – that between how well they could earn (relatively) in arbitrary jobs and workplaces.

[14]The increase between rows 4 and 5 may be due to the fact that students from schools whose students are systematically placed in low (high) wage occupations tend to achieve better (worse) wages than expected in the labour market.

Finally, we note that, even if we include tenth grade math scores in our decomposition, this indicator does not drop significantly. Therefore, it is evident that the observed patterns cannot be solely attributed to differences in measurable ability differences across schools.[15]

3.3 Differences across schools

In Table 3, we compare schools on the basis of the extreme values of some indicators, being interested in the difference in expected future wages between different schools and the channels through which these differences emerge. The first row of the table shows the total difference in log wages (which is why it will always be 100 percent), while the last row contains the systematic differences present at the school level that are not explained by anything else.[16]

Table 3: Contribution of each wage component to total wage differences between school types

	Raw wage Top - bottom quartile (1)		Math test score Top - bottom quartile (2)		School type General - vocational (3)	
	diff.	(%)	diff.	(%)	diff	(%)
Log wage	*0.366*	100.0	*0.281*	100.0	*0.154*	100.0
Gender and age	0.010	2.8	0.007	2.4	−0.002	−1.2
Tenure, region, year	0.036	9.8	0.039	13.7	0.026	17.1
Firm (obs. and unobs.)	0.173	47.4	0.093	32.9	0.028	18.4
Occupation (and sector)	0.077	21.0	0.084	30.1	0.071	46.1
Other unexplained	0.005	1.3	0.004	1.4	0.004	2.9
Educational attainment	0.011	3.1	0.013	4.8	0.010	6.8
10th grade math score	0.021	5.7	0.026	9.4	0.014	8.8
Unexplained individual	0.005	1.2	0.002	0.8	0.000	0.0
Adj. school effect	0.028	7.7	0.013	4.5	0.002	1.2

The first columns (comparison 1) show the average differences between schools in the top and bottom quartile of the ranking based on the raw future wages expected by their students (between the ages of 18 and 25). A pupil graduating from a better secondary school defined in this way can be expected to earn 0.366 log points (i.e. almost 44 percent) more. Almost half of the difference can be explained by being placed in a better firm (or industry), but selection into occupations also explains 21 percent. Only just under 9 percent of the differences can be explained by students from better schools achieving higher levels of education (3.1 percent) or having better observable (mathematical) skills (5.7 percent). Finally, around 7.7 percent can be still attributed to other systematic differences at school level. This again implies some indirect role of schools, beside cognitive skills.

When comparing schools in the upper and lower quartiles of the maths proficiency test (2) the findings closely mirror the previous comparison. The primary explanation for the variation is firm and occupational selection, with high school only accounting for a negligible 4.5 percent of the overall difference.

Finally, the analysis of high schools and vocational schools (3) shows that occupational selection clearly drives differences in this comparison. Not surprisingly, those with vocational education at the secondary level pursue different careers than those who have only completed high school, without pursuing tertiary education. Interestingly, we found no

[15](Narrowing the sample to students with scores reduces the fraction explained by total wage dispersion somewhat, presumably due to the non-random distribution of absenteeism on the day when the NABC is conducted.)

[16]That is we rely on the most detailed decomposition put forward in Equation 7.

significant difference in the last row, which suggests that both among upper secondary high schools and upper secondary vocational schools there are schools that perform better or worse in terms of future wages. Therefore, the fundamental difference between these two types of institutions is not the primary cause of the emergent wage disparities.

3.4 Sources of school heterogeneity (preliminary results)

As an attempt to understand the sources of school heterogeneity, we aim to decompose the variation in the estimated school effects according to Equation 17. Due to the limited availability of data, we use 8^{th} grade test scores instead of earlier proxies. Even this way, we face serious data attrition, as we lose 60% of observations we could use for this exercise. We control for completed education with four categories (university, secondary general, secondary vocational, only elementary) and for detailed school types (4, 6 or 8 grade general, vocational without and vocational with matriculation). First, we remove the variation coming from completed education, then decompose the remaining variance ($\text{var}(\xi_{s(i)} - \beta_H \text{H}_i)$). For comparison, first we do not include separate terms for selection and value added, just include 10^{th} grade test scores as a control. The results are presented in Table 4.

Table 4: Contribution of each wage component to total wage differences between school types

	(1)		(2)	
Component (C)	$\text{Cov}(\text{C}, \xi - \beta_H H)$	Contr. (%)	$\text{Cov}(\text{C}, \xi - \beta_H H)$	Contr. (%)
Adj. school quality	0.001066	100.0	0.000996	100.0
Test score	0.000067	6.3		
Selection			0.000056	5.6
Value added			0.000003	0.3
School type	0.000054	5.1	0.000039	3.9
Unexplained	0.000945	88.6	0.000898	90.2

We find that after removing the effect of further education from the school effects, of the remaining variation in adjusted school effects , around 90.2% of variation remains unexplained, 5.6% is captured by the 8^{th} grade test scores (pre-selection), 3.9% by school type, and only 0.3% relates to our proposed measure of (differences in) value added.[17] This would be, in some sense a negative result, implying no (additional) effect of high schools in explaining wage outcomes besides the student composition. We have to note however, that this figure only relates to the adjusted school effect, in which higher education, occupational choice, and firm selection are already accounted for. With these channels included, probably we would get higher importance of schools, as the differences between the two specifications in Table 2 suggests.

Interpreting the unexplained part as only the effect of non-cognitive skills would probably be also an over-interpretation. With respect to the current data limitations, we would avoid making strong claims based on these results. It would be, however, rather interesting to look at these results using the next iteration of the Hungarian LEE Data, Admin4. The dataset, that probably becomes available in 2024, will include data from five additional years, which would allow us to either use data from five more cohorts or also to look at outcomes at a slightly higher age. Both changes would greatly improve the validity of our results.

[17] In specification 1, we only include 10^{th} grade scores, and therefore we can use all 1 million wage observations for which a school effect has been estimated. However, there we can only find the total contribution of skills at the age of 16.

4 Discussion

Given the results presented in this study, what can we say about the importance of secondary education in determining future earnings inequality? Although a 4-10% contribution may seem small, international examples reveal it is a considerable impact. The fact that attending a particular secondary school at age 14 can determine up to 10% of a person's wage a decade later is striking. It is important to note that our model has not yet controlled for family background or childhood residence, which may significantly affect school choice and enrollment.[18] As a result, our estimate of the role of schools may be overestimated. Compared to international examples, the current study also examines a notably younger age group in the data, including early entrants with tertiary education (or, in some cohorts, not even including them). This could lead to overestimation of the average effects measured, considering the anticipated reduced role of secondary schooling for the higher educated who are less likely to be observed directly. However, this expectation that students are more likely to pursue further education from institutions with significant wage effects somewhat mitigates the aforementioned issue.

In order to improve the precision of our findings, both for the literature and for policy makers, it would be essential to include new cohorts, or more importantly, longer observation periods. Fortunately, this will be possible in the near future with the launch of the Admin4 database, which will include data up to the end of 2021 or 2022. Nonetheless, the present findings are already important, as they highlight that the variation among schools are not solely attributable to further education, occupational, or sectoral choices, but also have some direct explanatory power. Additionally, these disparities cannot be solely traced back to variations in cognitive skills, as indicated by scores, but also extend to a broader impact, most likely due to the cultivation of other, non-cognitive skills by better schools, or the various accumulation of social capital in certain schools. Nonetheless, exploring and ,most importantly, distinguishing between the exact mechanisms necessitates further study.

[18] Although the NABC includes a household survey, the fulfillment of the survey is not enforced. Hence various non-random sample selection issues would emerge if we would try to incorporate measures from the survey.

References

Abowd, J. M., Kramarz, F., & Margolis, D. N. (1999). High wage workers and high wage firms. *Econometrica*, *67*(2), 251–333. https://doi.org/10.1111/1468-0262.00020

Altonji, J. G., & Mansfield, R. K. (2011). The role of family, school, and community characteristics in inequality in education and labor-market outcomes. *Whither opportunity?: Rising inequality, schools, and children's life chances.*

Boza, I. (2021). Wage Structure and Inequality : The role of observed and unobserved heterogeneity. *KRTK-KTI WORKING PAPERS*, (31).

Boza, I. (2022). *Wage Differentials and the Possibilities in Linked Employer-Employee Panel Data* (Doctoral dissertation). Central European University. https://doi.org/10.14754/CEU.2022.01

Boza, I., & Ilyés, V. (2020). Decomposition of co-worker wage gains. *IZA Journal of Labor Economics*, *9*(1). https://doi.org/10.2478/izajole-2020-0008

Card, D., Cardoso, A. R., Heining, J., & Kline, P. (2018). Firms and Labor Market Inequality: Evidence and Some Theory. *Journal of Labor Economics*, *36*(S1), S13–S70. https://doi.org/10.1086/694153

Cardoso, A. R., Guimarães, P., & Portugal, P. (2016). What drives the gender wage gap? A look at the role of firm and job-title heterogeneity. *Oxford Economic Papers*, *68*(2), 506–524. https://doi.org/10.1093/oep/gpv069

Cardoso, A. R., Guimarães, P., Portugal, P., & Reis, H. (2018). The Returns to Schooling Unveiled. *IZA Discussion Papers*, (11419), 1–45. https://ideas.repec.org/p/iza/izadps/dp11419.html

Hermann, Z., Horn, D., Köllő, J., Sebők, A., Semjén, A., & Varga, J. (2022). A kompetenciaeredmények hatása a munkaerőpiaci sikerességre. *Közgazdasági Szemle*, *69*(2), 177–198. https://doi.org/10.18414/KSZ.2022.2.177

Hermansen, A. S., Borgen, N. T., & Mastekaasa, A. (2020). Long-Term trends in adult socioeconomic resemblance between former schoolmates and neighbouring children. *European Sociological Review*, *36*(3), 366–380. https://doi.org/10.1093/esr/jcz066

Ilyés, V., & Sebők, A. (2020). Egyetemről a munkaerőpiacra. Felsőoktatási ismeretségek hatása a munkaerőpiaci kilátásokra. *Közgazdasági Szemle*, *67*(10), 993–1028. https://doi.org/10.18414/KSZ.2020.10.993

Lindahl, L. (2011). A comparison of family and neighborhood effects on grades, test scores, educational attainment and income-evidence from Sweden. *Journal of Economic Inequality*, *9*(2), 207–226. https://doi.org/10.1007/S10888-010-9144-1/METRICS

Raaum, O., Salvanes, K. G., & Sørensen, E. (2006). The neighbourhood is not what it used to be. *Economic Journal*, *116*(508). https://doi.org/10.1111/j.1468-0297.2006.01053.x

Torres, S., Portugal, P., Addison, J. T., & Guimarães, P. (2018). The sources of wage variation and the direction of assortative matching: Evidence from a three-way high-dimensional fixed effects regression model. *Labour Economics*, *54*, 47–60. https://doi.org/10.1016/j.labeco.2018.06.004