Shakespeare vs. Marx vs. Einstein: Which Type of Education Offers the Highest Return?

PRELIMINARY AND INCOMPLETE VERSION - DO NOT QUOTE

CARL-JOHAN DALGAARD
Department of Economics, University of Copenhagen, and Centre for Economic and Business Research (CEBR),

MICHAEL ROSHOLM
Department of Economics, Aarhus School of Business and Centre for Economic and Business Research (CEBR),

ESBEN ANTON SCHULTZ
Department of Economics, Copenhagen Business School, and Centre for Economic and Business Research (CEBR),

AND

ANDERS SØRENSEN
Department of Economics, Copenhagen Business School, and Centre for Economic and Business Research (CEBR).

May 2007

Abstract

This paper investigates the importance of the field of study for the returns to education. Since (unobservable) factors are likely to influence the choice of educational type, an IV approach is applied. The main instrument is based on the hypothesis that peer effects may cause individuals to choose another field of study than the one consistent with their cognitive abilities. More specifically, we assume that the choice of educational type of a high school cohort is affected by the previous cohort's educational choices. This hypothesis is strongly supported. We find high variation in the returns to educational types with natural sciences having the lowest return and medical sciences the highest. Surprisingly, we do not find that human arts perform the worst. The IV estimates differs substantially from the OLS counterparts.

1 Introduction

The notion that education is a key determinant of individual productivity has a long and distinguished history in economics, going back (at least) to the work of Mincer (1958), Houthakker (1959) and Miller (1960). At the conceptual level one may distinguish between three dimensions of a formal education with holds the potential to affect individual productivity: The quantity of education, the quality of education and the subject matter studied.

While the quantity of education can be measured by years of schooling, the quality of education is harder to account for. Still, one may attempt to gauge the impact from quality, by adding reasonable proxies to otherwise standard wage regressions, such as test score results. Alternatively, one may try to infer the impact from quality by including characteristics of the school attended in earnings regressions (e.g. pupil/teacher ratios and school size). As is well known, standard theories would predict a positive impact from both of these dimensions of education on individual productivity (Becker, 1967), as well as on macroeconomic outcomes (e.g. Lucas, 1988). This proposition has been tested (and debated) intensely over the years.[1]

The third dimension of human capital accumulation has received considerably less attention by academic researchers. The issue is whether the particular field of study, or the contents of the curriculum, has a separate impact on individual productivity and ultimately macroeconomic performance.

From a macroeconomic perspective it is often argued that natural science is a field which should be prioritized by policy makers. The argument is based on the premise that a key driver of productivity growth is R&D. Since R&D activities chiefly are undertaken by individuals with a background in the natural sciences or engineering, one might argue that these fields should be supported in particular ways (e.g. Romer, 2000).[2]

From a microeconomic perspective one might similarly hypothesize that

[1] See Card (1999, 2001) for a review of the literature which attempts to estimate the causal impact from an additional year of schooling on individual wages; Card and Krueger (1996) review the literature on the impact from school quality on labor market outcomes at the level of the individual. Bills and Klenow (2000) provides an analysis of the education/growth nexus at the aggregate level; Hendricks (2002) examines the contribution from quality differences in human capital in accounting for cross-country wage differences.

[2] Some empirical macro studies can be seen as broadly supportive of this view. Hanushek and Kimko (2000) find a large impact of test scores in mathematics and science on economic growth, and Benavot (1992) find a positive growth effect from a greater science content in primary schooling, while no impact from higher emphasis on languages or arts and music. While suggestive, the issue of causality remains. Moreover, Hanushek and Kimko (2000) themselves view their findings as an indication of a positive impact from quality of education, rather than as support in favor of policies aimed at prioritizing science and mathematics in the curriculum.

some types of knowledge have a larger labor market pay-off compared with others. A few studies have adressed this question: Bishop (1992) and Joensen and Nielsen (2006) find that greater skills in mathematics goes along with higher wages, and Dougherty (2003) find that numeracy has a strong positive impact on individual wages, whereas literacy has a much smaller (and often insignificant) impact. These findings could be taken to suggest that skills attainable through studies within the broad fields of mathematics and the natural sciences are more valuable in the labor market than skills attainable within the humanities.[3]

The present paper contributes to this line of inquiry, by attempting to elicit information about the causal effect of the field of study on individual productivity, as it manifests itself in individual wages. The data set underlying the empirical analysis covers the part of the Danish population which completed high school during the period 1981-1990.[4] Narrowing the focus to the group of individuals which subsequently proceeded to a tertiary level of education (roughly 43,000), we ask whether wage rates differ systematically between individuals who chose to follow different fields of study. Specifically, we examine the relative labor market performance of individuals who chose to study within the broad fields of natural sciences, technical sciences, medical sciences, human arts and social sciences. Conditional on standard determinats of wages (such as experience), an ordinary least squares (OLS) regression reveals that individuals who pursued an education within the human arts fared the worst, while students of medical science did best.

However, OLS estimates are unlikely to capture the causal effect of the type of education on individual productivity. The central concern is that (unobservable) factors that affect individual productivity might also impact on the choice of which field to study. Suppose, for example, that skills in mathematics raise wages for a given amount of education (as Bishop (1992) finds). Suppose further, as seems plausible, that individuals who chose to study human arts were among the least mathematically predispositioned in the group pursuing a tertiary education. In that case one would expect OLS to underestimate the true impact from an education within the humanities and overestimate the impact from taking an education within the natural sciences, unless "math skills" are controlled for in the regression. Hence, this precise problem can be dealt with

[3] Indeed, a theoretical literature aims to explain why individuals would chose to pursue a type of education which holds a relatively low economic return. The general approach consists of arguing in favor of differences across lines of study in ways of pure utility (i.e. schooling as consumption), or differences in terms of the difficulty in pursuing various fields. See e.g. Alstadsæter et al. (2005) and Malchow-Møller and Skaksen (2003).

[4] When we refer to the Danish high school in this paper, we mean the ordinary high school ("gymnasium").

by including information about math skills in the wage regression. However, it is hard to rule out that other – unobserved – factors could simultanously impact on the choice of education type as well as productivity. Consequently, we also examine the issue by employing an instrumental variables (IV) approach.

Our identification strategy is inspired by the literature on peer effects. Accordingly, we hypothesize that peer effects may cause individuals to potentially choose another line of study than the one consistent with their cognitive abilities. The potential importance of peer effects is documented by Sacerdote (2001) who find that (randomly assigned) college roommates strongly influence each other in the context of academic achievements, and in decisions to join social groups. In our context we hypothesize that similar group dynamics may matter, for abilities and intellectual predispositions given, for the type of education the individual ends up pursuing. More specifically, we examine whether the choice of educational type of a high school cohort is affected by the educational choices of the *previous* cohort. We find strong support for such an association, which is consistent with peer effects running from the older kids to the younger ones. Admittedly, this correlation could in theory also be caused by other high school specific characteristics. For example, it could reflect the influence from the high school teaching staff within particular fields of study (and its quality). To check the exclusion restriction we therefore introduce an additional candidate instrument.

The fundamental idea behind the second instrument is that students may be influenced by the environment within which they grew up. That is, by factors outside the high school. Specifically, we hypothesize that the population share of various education types at the tertiary level, in the area where the individual grew up, may influence educational choices. Hence, as an instrument for the probability of choosing human arts, at the tertiary level, we also use the population share of students of the human arts in the local area where the person in question grew up. Likewise, if the individual lives in an area where an unusually large fraction of population falls within the natural sciences we expect a slightly higher probability of ultimately choosing to pursue natural science at the tertiary level, and so on.[5] These are all conditional statements; i.e., in the first stage we also control for individual specific abilities (proxied by over-all grade point average), and academic interests in a narrower sense (human arts, natural science etc.), as explained below.

The virtue of introducing a second instrument is that it allows us to test the exclusion restriction. That is, we can test whether our instruments can be

[5] The local area is defined as the political entity of a municipality (Danish,"kommune").

excluded from the second stage: The wage equation. Data does not allow us to reject this restriction.[6]

Our IV estimates differs substantially from the OLS counterparts. While the return to pursuing an education within medical sciences is still the highest, human arts no longer perform the worst. Instead, pure students of natural sciences have the lowest return. Moreover, the return "gap" between technical sciences and human arts shrinks. These findings suggest that the main reason why human arts students fare worse in the labor market, than natural science students, is the selection mechanism. The wage difference is not attributable to a lower inherent return on taking a human arts education.

The paper proceeds as follows. The next section provides a brief review of related literature (still to be completed). Section 3 presents a simple model illustrating the basic identification problem in our empirical analysis. Section 4 briefly describes the institutional settings of the tertiary educational system in Denmark. Section 5 describes the data set used in our empirical analysis. Section 6 presents our empirical strategy, while section 7 presents our main results. Finally, section 8 concludes.

2 Related Literature

To be added.... Kalmijn and van der Lippe (1997); Bratti and Mancini (2003) and O'Leary and Sloane (2005); Jacobsen et al. (2004).

3 Some Theoretical Considerations

The following simple model illustrates the identification problem. Consider an individual who plans on attending school, by studying two topics: H (human arts, say) and M (math). The number of years spent on topic H is S_H, while years used studying the M-subject is denoted by S_M.

Preferences of the individual is defined over earnings (y) and schooling

$$\log(y) - \left(c_H S_H + \frac{1}{2} S_H^2\right) - \left(c_M S_M + \frac{1}{2} S_H^2\right), \tag{1}$$

where c_i is the perceived costs of schooling within topic i=H, M, capturing opportunity costs, tuition fees, as well as psychic costs of schooling. The latter could be negative; utility from studying topic i.

[6] A counter argument with respect to the exogeniety of the second candidate instrument would be that the educational type is partly correlated with ability, and that high ability individuals tends to end up living in the same area. By focusing only on individuals with a tertiary education we believe that this selection mechanism is unlikely to be important in practise. In any case, we cannot reject the exclusion restriction.

A likely source of heterogeniety with respect to the c's could be differences in relative abilities in studying the two topics. Let a denote the ability of an individual in studying M, and assume that

$$c_M = c\left(a; \mathbf{x}\right),$$

where $\partial c_M / \partial a < 0$. Hence, a more "talented" person will face lower costs of pursuing topic M. Note, that by keeping c_H independent of a, we implicitly allow a to capture both *absolute* advantages in intellectual pursuits ("IQ") and *relative* talent for the M-topic.[7] Now, in addition to a, we also assume that other factors, unrelated to inherent abilities and tastes for subjects, could matter for the perceived costs – e.g. peer effects. Such factors are denoted by $\mathbf{x}$.

Earnings are given by

$$y = h \cdot e^{\epsilon} \tag{2}$$

where h reflect the skill level of the individual, and ϵ represent factors which affect wages aside from human capital.

Finally, we assume skills are produced with the following technology:

$$h = e^{\beta(S_M + S_H)} e^{a}. \tag{3}$$

Accordingly, for education given high-ability individuals will have higher stocks of human capital than less able individuals. Note that we assume time spent on either subject is equally productive in building up human capital: One extra year of training leads to e^{β} units of skills, no matter whether time is spend on subject H or M.

Solving the problem of maximizing (1) subject to (2) and (3) yields

$$S_M = \beta - c_M\left(a; \mathbf{x}\right) \wedge S_H = \beta - c_H. \tag{4}$$

Against this background consider the following earnings regression

$$\log\left(y\right) = \beta S_M + \beta S_H + a + \epsilon, \tag{5}$$

where the levels of schooling is given by equation (4).

Estimating equation (5) by OLS will obviously only provide us with the true return to schooling, within the two areas, if a can be fully accounted for. If some dimensions of a are unobservable and thus left in the residual, the return

[7] See Willis (1986) for discussion of human capital as a multidimensional concept involving inherent comparative advantages in various "tasks". Alternatively, see Alstadsæter et al (2005) for an analysis where general abilities makes it less costly to pursue an education within technical lines of study.

estimates will be biased since $\partial S_M/\partial a > 0$ (cf. eq (4)); the return on time spend studying M will be biased in an upward direction,

Section 6 explains how we try to adress this fundamental problem. Essentially we approach the matter in two ways. First, we attempt to control carefully for a. That is, we try to capture both over-all abilities (e.g. by including grade point averages), *and* relative talents in intellectural pursuit of various topics. Second, we look for variables which matter for the perceived relative costs of pursuing either topic H or M, while at the same time are uncorrelated with a. In short, we are looking for elements in $\mathbf{x}$ to use as instruments for the type of education.

4 Institutional Settings

In this paper we focus exclusively on individuals who have completed a tertiary education, since we want to avoid any selection bias in our results due to the choice of length of education. In this section we briefly describe some of the basic institutional features in relation to the tertiary educational system in Denmark.

In contrast to the Anglo-Saxon system, there are no tuition fees for education in Denmark. In fact, all Danish students above 18 receive a monthly study grant from the government that suffices to cover living expenses and do not depend on financial circumstances. However, to be enrolled in any institution of tertiary education applicants still have to meet some basic admission requirements. As a general admission requirement applicants must have completed high school.[8] At the same time, each institution has specific requirements on the high school grade point average. Furthermore, some institutions also require that applicants have taken specific high school courses. Applicants who do not meet these requirements can, however, still be taken into consideration in the enrolment process if they - in addition to their high school achievements - can provide evidence of other relevant qualifications, such as labor market experience, staying abroad, additional education and so on.

Individuals can enter the Danish high school immediately after completing lower secondary school and it is prescribed to take three years to complete. Since the beginning of the 1960's, the Danish high school system has been characterized by a distinction between an education based on a mathematical track or on a language track. At the same time, the Danish high school were

[8] Only in very few cases, applicants without a high school degree can be enrolled at a higher education. However, as these few individuals in general tend to have significantly different characteristics (such as age, labour market experience, etc.) than applicants with a high school degree, we ignore this group in our analysis.

until 1988 based on a branch-based regime, where courses were grouped in strictly defined course packages. However, with the structural reform of the Danish high school system in 1988, the branch-based regime was replaced by a choice-based regime with combinations of optional courses on two levels, which students themselves could combine within the given framework. In this paper our focus is on students who graduated from the branch-based regime, since it provide us with precise information of what courses students took and the level of each course. A further advantage of this focus is that our data set includes a longer period of labour market outcomes of these individuals.

In the branch-based regime, students had to choose between the math track and the language track upon entry to high school. After their first year of high school, they then had to choose one of the branches available for each track. Math track students could choose between Math/Physics, Math/Natural Sciences, Math/Social Sciences or Math/Music, while languages track students could choose between Languages/Social Sciences, Languages/Music, Modern Languages or Classical Languages.[9] As discussed in section 3, our instrument to correct for endogeneity of the choice of educational types is based on these eight different branches.

5 Data

The data we use in our empirical analysis is a very rich panel data set covering the danish population of individuals completing high school in the years 1978-1990 in Denmark. The data are administered and maintained by Statistics Denmark, who has gathered the data from three administrative registers: the Integrated Database for Labor Market Research (IDA), the Danish Income Registry and the Danish Student Registry.

For each individual, we have complete data on educational and labor market histories along with detailed information on other socio economic characteristics. The educational data comprise detailed codes for the type of education attended (level, subject, and educational institution) and the year for completing the education. Furthermore, we have information on the branch of choice in high school and the high school grade point average.[10] The grade point av-

[9] At some high schools they also implemented a fifth branch as a pilot scheme for math track students and language track students, respectively. However, as only very few students graduated from these branches, we have treated them as Math/Music students and Language/Music students, respectively.

[10] In the period we analyze, a numerical grading scale system is used in Denmark. The possible grades are 00, 03, 5, 6, 7, 8, 9, 10, 11 and 13, where 6 is the lowest passing grade, and 8 is given for the average performance.

erage is a weighted average of the grades at the final exam at each course. The quality of the courses as well as the grade point average is comparable across high schools since the control of the high school is centralized at the Danish Ministry of Education. In addition, all students within each high school cohort faces identical written exams, and the oral exams and the major written assignments are evaluated both by the student's own teacher and an external examiner assigned by the Ministry of Education.

We have precise information on hourly wage. This variable is calculated as the annual labor income divided by hours worked. Other individual background variables that we use in our estimations are gender, actual labor market experience, municipality, county and sector of occupation (in later versions we will use more controls).

Among the gross population of high school students in the years 1978-1990, we select only high school graduates who subsequently complete a tertiary education.[11] We condition on them being wage earners for at least one year, that is, to have a positive income from wage employment for at least one year. We further restrict the sample to individuals aged above 23 in a given year. After these restrictions, there are 53,257 individuals in our sample. As we shall explain below, we use past cohorts to instrument for educational choice, and as additional instuments we use the fractions of adults in the individual's municipality of residence at the time of high school completion with different types of education. This latter information can only be calculated from 1981. Hence, high school graduates from 1978-1980 are removed from the sample. This leaves 42,857 individuals, who constitute our base sample. They come from 151 different high schools.

In the wage equations estimated below, these individuals are followed from the time of their graduation (earliest graduation year is 1985) from the educational system and until 2003. We condition on them being wage earners, that is, having positive income wage income. This produces a sample of 387,145 observations.

5.1 Descriptive Statistics

Table 1 displays descriptive statistics for all high school related variables. The sampling unit here is the individual, and the table presents the distribution of students on high school branches, their high school grade point average, their gender and their county. In the appendix, we provide the distribution of gender,

[11] This corresponds to about 30% of the gross population of high school students in the years 1978-1990.

grade etc. by high school branch (still to be completed). The table also presents the distribution on subsequent type of tertiary education. Social sciences comprise predominantly lawyers, economists and political scientists. Human arts consist, for instance, of linguistics, theologians and historians. Medical sciences consist primarily of doctors and dentist, while technical sciences comprise mainly various types of engineers. Finally, natural sciences cover e.g. biologists, physicists and chemists. In the appendix, we provide a detailed definition of each type of tertiary education (still to be completed).

Table 1. Descriptive statistics for high school graduates who subsequently obtain a tertiary education

	Mean	Std.dev.
Subsequent education type		
Natural sciences	0.111	
Medical sciences	0.108	
Technical sciences	0.246	
Human arts	0.164	
Social sciences	0.371	
High school branch		
Math/Music	0.075	
Math/Physics	0.337	
Math/Natural Sciences	0.187	
Math/Social Sciences	0.154	
Modern Languages	0.109	
Classical Languages	0.010	
Languages/Social Sciences	0.094	
Languages/Music	0.034	
High school grade point average	8.840	0.864
Woman	0.537	
County of residence		
Copenhagen & Frederiksberg Municipalities	0.366	
Copenhagen county	0.153	
Frederiksborg	0.074	
Roskilde	0.037	
Western Sealand	0.020	
Storstroem	0.013	
Bornholm	0.002	
Funen	0.050	
Souther Jutland	0.015	
Ribe	0.013	
Vejle	0.030	
Ringkoebing	0.018	
Aarhus	0.135	
Viborg	0.017	
Northern Jutland	0.059	
Number of persons	42,857	

The descriptive statistics reveals some interesting insights. First, it is seen that around 75% choose the math track, while only 25% choose the language track. Furthermore, it is seen that the largest high school branch is Math/Physics. As regards subsequent educational choices, 36% choose social sciences, while only 17% choose humans arts, 25% choose technical sciences, 12% medical sciences and 11% choose natural sciences.

6 Econometric Strategy

The empirical counterpart to the model specified above is a wage equation that includes educational types. However, as mentioned above, these are likely to be endogenous. In the following, we describe the approach we employ in order to address this issue.

Let the log wage earned by individual i at time t be denoted $\log(y_{it})$. $\mathbf{s}_i$ denotes the type of education chosen, given that it is a tertiary education. That is, $\mathbf{s}_i$ is a set of indicators for having a masters degree in natural sciences, medical sciences, technical sciences, human arts, and social sciences. $\mathbf{z}_{it}$ is a set of observed background variables including e.g. gender and working experience. A naive wage equation may now be specified as

$$\log(y_{it}) = \alpha + \mathbf{z}_{it}\beta + \mathbf{s}_i\rho + \epsilon_{it}$$

where the error component captures unobserved factors. However, the educational choice may be determined in part by the individual's intelligence, both in terms of level and composition. Denote by a_i the (absolute) intelligence level of the individual, and by $\boldsymbol{\tau}_i$ the vector describing the relative talent of the individual.[12] Now, if the choice of education depends on the level of intelligence itself, then OLS is biased, because

$$\rho_{OLS} = \rho + \frac{\mathbf{E}\left[a_i\mathbf{s}_i\right]}{Cov(\mathbf{s}_i)}.$$

However, conditioning on the length of education, as we do in this paper, and including a specific proxy for the level of intelligence, namely, the grade point average obtained in high school, we expect this bias to be negligible.

The potentially important bias left, then, is the one resulting from correlation between the costs of choosing a given type of educaton and $\boldsymbol{\tau}_i$, i.e.

$$\rho_{OLS} = \rho + \frac{\mathbf{E}\left[\boldsymbol{\tau}_i\mathbf{s}_i\right]}{Cov(\mathbf{s}_i)}.$$

It is quite likely that "comparative advantages" across fields could simulatanously affect productivity (wages) and the choice of schooling *type*.

In order to remove this bias, we do two things; first, we include the choice of branch chosen in high school, which to some extent we expect to reflect your

[12] In contrast to Section 3 we now *distinguish* between over-all ability, and relative talent across fields of study. Hence, two individuals may have the same intelligence level, a, while having it "composed" in different ways. That is, the first being an exceptionally gifted linguist (say), but more ordinary when it comes to mathematical reasoning, and vice versa for the second individual.

own assessment of your costs of acquiring specific skill types. However, we expect it to be an imperfect proxy, as your choice of high school branch may also be determined by other factors as well. Therefore, we include instruments in terms of the peer effects discussed in section 3, that is, we construct a set of variables measuring the fraction of students in the *previous* cohort in your branch within your own high school taking each type of education. Now, presumably all individuals choose high school branch partly on the basis of their perceived abilities (costs), and therefore, these fractions measured within high school branch may also reflect cognitive abilities within the specific area of the high school branch. Therefore, we measure them for each high school branch within each high school as the deviation from the national fractions within each high school branch. Thus, the instrument measures high school specific deviations from the national education choices within each high school branch.

The instrument is constructed by estimating a multinomial logit for the choice of educational type (natural sciences, medical sciences, technical sciences, human arts, and social sciences), conditional on subsequently taking a tertiary education. Here, we include all relevant background variables measured at the time of high school graduation (years of actual work experience is not relevant to include in this equation, as this is zero for nearly all high school graduates at the time of graduation), high school grade point average, high school branch, and the variable measuring peer educational choice as specified above. The parameters from the logit model are then used to predict, for each high school graduate who proceeds to taking a tertiary education, the probability of taking each of the five types of tertiary education.

In the next step, we analyse the log wages of these individuals upon graduation from tertiary education. Here, we include as explanatory variables everything included in the logit model except the instruments, plus years of actual work experience, plus a set of indicators for calendar year (in order to correct for cyclical effects and inflation). Moreover, we include the predicted probability of making each educational choice, leaving out one (social science). The coefficients on these probabilities thus constitute unbiased estimates of the return to a given type of tertiary education.

7 Results

In the following, we first report the results from the estimation of choice of educational direction, and subsequetly, we report the results from the estimation of the returns to educational directions.

7.1 Educational choice

Table 2 reports the results from a multinomial logit model of choice of educational direction. The table reports marginal effects and their associated standard errors. The variables included in the model are gender and high school grade point average, high school branch, high school graduation year, and the instruments discussed above, the fraction of students in your high school and high school branch, graduating the year before you, who chose different educational directions, measured in deviations from these fractions measured at the national level within each high school branch. Note that one of the instruments is left out, as they sum to 0 (since they are measured in deviations from the national mean).

Table 2. Marginal effects from multinomial logit for choice of tertiary educational direction

	Natural sciences		Medical sciences		Technical sciences		Human arts		Social sciences	
	Marg. Eff.	Std. Err.	Marg. Eff.	Std. Err.	Marg. Eff.	Std. Err.	Marg. Eff.	Std. Err.	Marg. Eff.	Std. Err.
Woman	**-0.008**	**0.003**	**0.082**	**0.003**	**-0.074**	**-0.005**	**0.049**	**0.004**	**-0.050**	**0.006**
High school grade point average	**0.019**	**0.002**	**0.057**	**0.002**	0.001	0.003	**-0.022**	**-0.002**	**-0.056**	**0.003**
High school branch										
Math/Music	**-0.024**	**0.005**	-0.008	0.005	**-0.123**	**-0.005**	**0.207**	**0.014**	**-0.052**	**0.012**
Math/Physics										
Math/Natural Sciences	**0.041**	**0.005**	**0.076**	**0.005**	**-0.096**	**-0.005**	**0.047**	**0.008**	**-0.068**	**0.009**
Math/Social Sciences	**-0.098**	**-0.003**	**-0.082**	**-0.003**	**-0.248**	**-0.004**	**0.097**	**0.009**	**0.330**	**0.009**
Modern Languages	**-0.117**	**-0.002**	**-0.072**	**-0.002**	**-0.249**	**-0.003**	**0.480**	**0.012**	**-0.042**	**0.011**
Classical Languages	**-0.106**	**-0.003**	**-0.079**	**-0.003**	**-0.218**	**-0.004**	**0.621**	**0.021**	**-0.218**	**0.021**
Languages/Social Sciences	**-0.106**	**-0.003**	**-0.086**	**-0.002**	**-0.249**	**-0.003**	**0.303**	**0.012**	**0.138**	**0.012**
Languages/Music	**-0.104**	**-0.003**	**-0.080**	**-0.003**	**-0.217**	**-0.004**	**0.634**	**0.013**	**-0.234**	**0.012**
High school graduation year										
1981	-0.010	0.007	**0.079**	**0.010**	0.016	0.011	**-0.041**	**0.006**	**-0.045**	**0.013**
1982	**-0.020**	**0.006**	**0.077**	**0.009**	-0.012	0.010	**-0.042**	**0.006**	-0.004	0.013
1983	-0.011	0.006	**0.046**	**0.008**	-0.002	0.010	**-0.044**	**0.006**	0.011	0.012
1984	**-0.017**	**0.006**	**0.026**	**0.007**	-0.005	0.009	**-0.044**	**0.006**	**0.040**	**0.012**
1985	**-0.029**	**0.005**	0.010	0.007	0.002	0.009	**-0.039**	**0.006**	**0.056**	**0.012**
1986	**-0.022**	**0.006**	**-0.014**	**0.006**	0.002	0.010	**-0.025**	**0.006**	**0.058**	**0.012**
1987	**-0.014**	**0.006**	-0.010	0.006	0.001	0.010	**-0.029**	**0.006**	**0.052**	**0.012**
1988	-0.011	0.006	-0.006	0.006	-0.015	0.009	**-0.018**	**0.006**	**0.050**	**0.012**
1989	-0.002	0.006	-0.006	0.006	-0.006	0.009	-0.006	0.006	0.020	0.011
1990 (REF)										
County of residence										
Copenhagen & Frederiksberg Munipalities (REF)										
Copenhagen county	**-0.020**	**0.004**	**0.014**	**0.005**	**0.081**	**0.007**	**-0.061**	**-0.004**	-0.015	0.008
Frederiksborg	-0.008	0.006	**0.018**	**0.006**	**0.103**	**0.010**	**-0.056**	**-0.005**	**-0.057**	**0.011**
Roskilde	-0.001	0.008	**0.026**	**0.009**	**0.066**	**0.013**	**-0.044**	**0.007**	**-0.048**	**0.015**
Western Sealand	-0.011	0.010	**0.053**	**0.013**	**0.139**	**0.019**	-0.021	0.011	**-0.159**	**0.018**
Storstroem	-0.005	0.014	**0.048**	**0.016**	**0.104**	**0.024**	-0.003	0.015	**-0.145**	**0.023**
Bornholm	**-0.062**	**0.023**	0.049	0.042	0.102	0.061	0.022	0.040	**-0.111**	**0.058**
Funen	**0.022**	**0.008**	**0.118**	**0.011**	**-0.070**	**0.010**	**0.074**	**0.010**	**-0.144**	**0.012**
Souther Jutland	**-0.025**	**0.011**	**0.077**	**0.017**	**0.088**	**0.021**	-0.005	0.013	**-0.135**	**0.021**
Ribe	**-0.033**	**0.011**	**0.077**	**0.018**	**0.079**	**0.022**	-0.018	0.013	**-0.105**	**0.023**
Vejle	**-0.030**	**0.008**	**0.088**	**0.013**	**0.040**	**0.015**	-0.009	0.010	**-0.088**	**0.016**
Ringkoebing	**-0.030**	**0.010**	**0.065**	**0.016**	**0.107**	**0.021**	0.008	0.014	**-0.150**	**0.020**
Aarhus	**0.038**	**0.006**	**0.053**	**0.006**	**-0.043**	**0.007**	**0.015**	**0.006**	**-0.062**	**0.009**
Viborg	-0.006	0.012	0.026	0.013	**0.171**	**0.022**	**-0.031**	**0.011**	**-0.159**	**0.020**
Northern Jutland	**-0.057**	**-0.005**	**0.016**	**0.008**	**0.178**	**0.012**	-0.013	0.007	**-0.124**	**0.012**
Instruments										
Fraction of previous cohort choosing										
Natural sciences	**0.042**	**0.012**	-0.002	0.011	0.022	0.019	-0.010	0.015	**-0.052**	**0.024**
Medical sciences	**0.026**	**0.012**	-0.006	0.011	0.027	0.019	0.010	0.013	**-0.057**	**0.022**
Technical sciences	0.010	0.010	0.008	0.009	**0.099**	**0.015**	-0.009	0.011	**-0.108**	**0.018**
Human arts	0.004	0.012	-0.002	0.010	**0.044**	**0.017**	**0.041**	**0.008**	**-0.086**	**0.018**
Fraction of employed workers in Municipality with										
Natural sciences	**0.247**	**0.075**	**-0.142**	**0.069**	0.003	0.109	**0.224**	**0.083**	**-0.333**	**0.137**
Nedical sciences	0.010	0.033	**-0.090**	**0.030**	**0.119**	**0.047**	**0.163**	**0.036**	**-0.201**	**0.058**
Technical sciences	0.059	0.032	**-0.124**	**0.029**	**0.264**	**0.045**	0.051	0.035	**-0.249**	**0.057**
Human arts	**0.153**	**0.037**	**-0.122**	**0.033**	**0.226**	**0.053**	**0.181**	**0.040**	**-0.438**	**0.066**

Note: Bold numbers are statistically significant at the 5% level

7.2 Wage regressions

In Table 3, we report the results from a set of wage regression performed on the students whose educational choices are modelled above. In the first model (raw log wage differences), only the indicators of educational choice are included, to give an idea of the raw log wage differences between different actual education directions in terms of wages. In Models 2-3, we subsequently introduce more information into the log wage regression to investigate how the estimated returns to educational types change. Finally, in Model 4, we introduce predicted educational choice using only peer effects as exogenous instruments, rather than the actual choices, which were used in Models 1-5. Model 5 uses only municipal edcuational distributions as exogenous instruments, while Model 6 uses both sets of instruments.

Table 3. OLS and IV log wage regressions

	Raw log wage diff.		Model 2		Model 3		Model 4		Model 5		Model 6	
	Coeff.	Std. Err.	Coeff.	Std. Err.	Coeff.	Std. Err.	Coeff.	Std. Err.	Coeff.	Std. Err.	Coeff.	Std. Err.
Type of education												
Natural Sciences	**-0.0637**	**0.0023**	**-0.0134**	**0.0022**	**-0.0289**	**0.0023**	**-0.3656**	**0.0147**	**-0.3344**	**0.0145**	**-0.3251**	**0.0142**
Medical Sciences	**0.0984**	**0.0022**	**0.1660**	**0.0021**	**0.1484**	**0.0023**	**0.3003**	**0.0094**	**0.3029**	**0.0094**	**0.3003**	**0.0093**
Technical Sciences	**0.0667**	**0.0016**	**0.0288**	**0.0015**	**0.0074**	**0.0017**	**-0.0238**	**0.0071**	-0.0054	0.0071	-0.0067	0.0070
Human arts	**-0.2635**	**0.0020**	**-0.1527**	**0.0020**	**-0.1444**	**0.0021**	**-0.1185**	**0.0090**	**-0.0849**	**0.0089**	**-0.0861**	**0.0087**
Social Sciences (REF)												
Working experience/10			**0.4624**	**0.0035**	**0.4563**	**0.0035**	**0.4747**	**0.0036**	**0.4761**	**0.0036**	**0.4763**	**0.0036**
Working experience squared/100			**-0.1447**	**0.0014**	**-0.1438**	**0.0014**	**-0.1495**	**0.0014**	**-0.1500**	**0.0014**	**-0.1500**	**0.0014**
Woman			**-0.1158**	**0.0013**	**-0.1053**	**0.0013**	**-0.1235**	**0.0016**	**-0.1233**	**0.0016**	**-0.1230**	**0.0016**
Private Sector			**0.1396**	**0.0013**	**0.1429**	**0.0013**	**0.1365**	**0.0013**	**0.1364**	**0.0013**	**0.1363**	**0.0013**
High school grade/10					**0.3356**	**0.0072**	**0.3108**	**0.0083**	**0.3124**	**0.0083**	**0.3120**	**0.0083**
High school branch												
Math/Music					**-0.0924**	**0.0027**	**-0.0974**	**0.0029**	**-0.0983**	**0.0029**	**-0.0983**	**0.0029**
Math/Physics (REF)												
Math/Natural Sciences					**-0.0611**	**0.0018**	**-0.0628**	**0.0022**	**-0.0624**	**0.0022**	**-0.0627**	**0.0022**
Math/Social Sciences					**-0.0423**	**0.0020**	**-0.0718**	**0.0031**	**-0.0653**	**0.0031**	**-0.0650**	**0.0031**
Modern Languages					**-0.0506**	**0.0024**	**-0.1001**	**0.0033**	**-0.1026**	**0.0033**	**-0.1017**	**0.0033**
Classical Languages					**-0.0683**	**0.0064**	**-0.1200**	**0.0071**	**-0.1262**	**0.0071**	**-0.1251**	**0.0071**
Languages/Social Sciences					**-0.0634**	**0.0025**	**-0.0994**	**0.0033**	**-0.0975**	**0.0033**	**-0.0970**	**0.0033**
Languages/Music					**-0.0961**	**0.0040**	**-0.1492**	**0.0050**	**-0.1565**	**0.0050**	**-0.1555**	**0.0050**
Geographic indicators	NO		YES		YES		YES		YES		YES	
Annual indicators	YES		YES		YES		YES		YES		YES	
# observations	387,145		387,145		387,145		387,145		387,145		387,145	
R-squared	0.237		0.338		0.346		0.329		0.329		0.329	

Note: Raw log wage differences has a set of annual indicators and the educational type of indicators. Model 2 adds gender, working experience and its square, and an indicator for working in the private sector. Model 3 adds high school grade point average and branch. Model 4 uses only the first set of instruments (previous cohort). Model 5 uses only the second set of instruments (municipality fractions, and Model 6 uses both sets of instruments.

Table 4. OLS and IV log wage regressions, private sector only

	Raw log wage diff.		Model 2		Model 3		Model 4		Model 5		Model 6	
	Coeff.	Std. Err.	Coeff.	Std. Err.	Coeff.	Std. Err.	Coeff.	Std. Err.	Coeff.	Std. Err.	Coeff.	Std. Err.
Type of education												
Natural Sciences	**-0.0370**	**0.0034**	**-0.0081**	**0.0032**	**-0.0279**	**0.0033**	**-0.3850**	**0.0206**	**-0.3434**	**0.0204**	**-0.3359**	**0.0200**
Medical Sciences	**0.0758**	**0.0038**	**0.1358**	**0.0036**	**0.1279**	**0.0038**	**0.3056**	**0.0139**	**0.3073**	**0.0138**	**0.3041**	**0.0138**
Technical Sciences	**0.0432**	**0.0020**	**0.0250**	**0.0020**	-0.0002	0.0022	**-0.0297**	**0.0097**	-0.0073	0.0097	-0.0096	0.0095
Human arts	**-0.3601**	**0.0030**	**-0.2211**	**0.0029**	**-0.2098**	**0.0031**	**-0.1646**	**0.0128**	**-0.1196**	**0.0126**	**-0.1240**	**0.0124**
Social Sciences (REF)												
Working experience/10			**0.5346**	**0.0048**	**0.5244**	**0.0048**	**0.5408**	**0.0049**	**0.5426**	**0.0049**	**0.5426**	**0.0049**
Working experience squared/100			**-0.1678**	**0.0019**	**-0.1656**	**0.0019**	**-0.1706**	**0.0019**	**-0.1711**	**0.0019**	**-0.1711**	**0.0019**
Woman			**-0.1510**	**0.0018**	**-0.1388**	**0.0019**	**-0.1630**	**0.0023**	**-0.1628**	**0.0023**	**-0.1624**	**0.0023**
High school grade/10					**0.3444**	**0.0100**	**0.2718**	**0.0115**	**0.2733**	**0.0115**	**0.2728**	**0.0115**
Highschool branch												
Math/Music					**-0.1030**	**0.0037**	**-0.1012**	**0.0039**	**-0.1027**	**0.0039**	**-0.1025**	**0.0039**
Math/Physics (REF)												
Math/Natural Sciences					**-0.0639**	**0.0025**	**-0.0641**	**0.0031**	**-0.0638**	**0.0031**	**-0.0642**	**0.0031**
Math/Social Sciences					**-0.0434**	**0.0027**	**-0.0672**	**0.0043**	**-0.0592**	**0.0043**	**-0.0592**	**0.0042**
Modern Languages					**-0.0541**	**0.0035**	**-0.1057**	**0.0048**	**-0.1097**	**0.0047**	**-0.1084**	**0.0047**
Classical Languages					**-0.1663**	**0.0111**	**-0.2099**	**0.0121**	**-0.2193**	**0.0120**	**-0.2172**	**0.0120**
Languages/Social Sciences					**-0.0781**	**0.0037**	**-0.1139**	**0.0047**	**-0.1119**	**0.0047**	**-0.1113**	**0.0047**
Languages/Music					**-0.1060**	**0.0059**	**-0.1654**	**0.0073**	**-0.1756**	**0.0073**	**-0.1737**	**0.0073**
Geographic indicators	NO		YES		YES		YES		YES		YES	
Annual indicators	YES		YES		YES		YES		YES		YES	
# observations	235,737		235,737		235,737		235,737		235,737		235,737	
R-squared	0.258		0.345		0.353		0.329		0.338		0.329	

Note: Raw log wage differences has a set of annual indicators and the educational type of indicators. Model 2 adds gender, working experience and its square, and an indicator for working in the private sector. Model 3 adds high school grade point average and branch. Model 4 uses only the first set of instruments (previous cohort). Model 5 uses only the second set of instruments (municipality fractions, and Model 6 uses both sets of instruments.

8 Conclusion

To be added.

References

[1] Alstadsæter A., A-S. Kolm and B. Larsen, 2005. Money or Joy: The Choice of Education Type. Working Paper (Copenhagen Business School)

[2] Becker, G., 1967 : Human Capital and the Personal Distribution of Income. Ann Arbor Michigan: University of Michigan Press.

[3] Benavot, A., 1992. Curricular Content, Educational Expansion, and Economic Growth. *Comparative Education Review*, 36, 150-174

[4] Bishop, J. H., 1992. The impact of academic competencies on wages, unemployment, and job performance. *Carnegie-Rochester Conference Series on Public Policy*, 37, 127–194.

[5] Bills M. and P. Klenow, 2000. Does Schooling Cause Growth? *American Economic Review*, 90. 1160-82

[6] Bratti M. and L. Mancini, 2003. Differences in Early Occupational Earning of UK Male Graduates by Degree Subject: Evidence from the 1980-93 USR. IZA Discussion Paper No. 890

[7] Card, D., 1999. The Causal Effect of Education on Earnings. In *Handbook of Labor Economics*, Ashenfelder O. and D. Card (eds), Vol. 3, Ch. 30. Elsevier.

[8] Card, D., 2001. Estimating the return to schooling: progress on some persistent econometric problems. *Econometrica,* 69, 1127-60

[9] Card, D. and A. Krueger, 1996. Labor Market Effects of School Quality: Theory and Evidence. NBER working paper 5450.

[10] Dougherty, C., 2003. Numeracy, literacy and earnings: evidence from the National Longitudinal Survey of Youth. *Economics of Education Review*, 22, 511-521

[11] Hanushek E.A. and D.D. Kimko, 2000. Schooling, Labor-Force Quality, and the Growth of Nations. *American Economic Review*, 90, 1184-1208.

[12] Hendricks, L., 2002. How Important Is Human Capital for Development? Evidence from Immigrant Earnings. *American Economic Review*, 92, p. 198-219.

[13] Houthakker, H.S., 1959. Education and Income. *Review of Economics and Statistics*, 41, 24-28.

[14] Jacobsen L.,R. LaLonde and D. Sullivan, 2004. Estimating the Returns to Community College Schooling for Displaced Workers. IZA Discussion Paper No. 1017.

[15] Joensen J. and H.S. Nielsen, 2006. Is there a Causal Effect of High School Math on Labor Market Outcomes? IZA Discussion Paper No. 2357.

[16] Kalmijn M. and T. van der Lippe, 1997. Type of Schooling and Sex Differences in Earnings in Netherlands. *European Sociological Review*, 13, 1-15.

[17] Lucas, R. Jr., 1988. On the Mechanics of Economic Development. *Journal of Monetary Economics*, 22, 3-42.

[18] Malchow-Møller, N. and J.R. Skaksen, 2003. How to Finance Education - Taxes or Tuision Fees. CEBR working paper 2003-28.

[19] Mincer, J., 1958. Investment in Human Capital and Personal Income Distribution. *Journal of Political Economy*, 66, p. 281-302

[20] Miller, H.P., 1960. Annual and Life Time Income in Relation to Education. *American Economic Review*, 50, 962-86.

[21] O'Leary N.C. and P.J. Sloane, 2005. The return to a university education in great britain. *National Institute Economic Reviem* 19, p. 75-89

[22] Romer, P., 2000. Should the government subsidize supply or demand in the market for scientists and engineers? NBER working paper 7723.

[23] Sacerdote, B., 2001. Peer Effects With Random Assignment: Results For Dartmouth Roommates, *Quarterly Journal of Economics*, 116, 681-704

[24] Willis, R.J., 1986. Wage Determinants: A Survey and Reinterpretation of human capital earnings functions. In *Handbook of Labor Economics,* Ashenfelter O. and R. Layard (eds), Vol. 1, Ch. 10. Elsevier.