Occupations, Fields of Study, and Returns to Education*

Thomas Lemieux, Vancouver School of Economics,
University of British Columbia

Revised, September 2014

Abstract

This paper considers several possible channels behind the well-documented effect of education on earnings. The first channel is that education makes workers more productive on a given task, as in a conventional human capital framework. The second channel is based on the idea that education helps workers get assigned to higher-paying occupations where output is more sensitive to skill. A third and final channel is that workers are more productive and earn more when they are matched to a job related to their field of study. Using data from the 2005 National Graduate Survey and the 2006 census, I find that the two latter channels account for close to half of the conventionally measured return to education. The results also indicate that the return to education varies greatly depending on occupation, field of study, and the match between these two factors.

*Presidential address delivered at the 2014 Annual Meetings of the Canadian Economic Association in Vancouver, British Columbia. I would like to thank Michael Baker, Dwayne Benjamin, and Nicole Fortin for useful comments, and SSHRC and the Bank of Canada Fellowship program for research support.

1. INTRODUCTION

One of the best established facts in economics is that the effect of education on earnings is large and robust across time and space. It has long been known that more educated workers earn substantially more than their less-educated peers. This empirical regularity was the foundation of the human capital approach to education and earnings pioneered by Schultz, Becker and Mincer. The famous Mincer earnings equation (Mincer, 1974) typically yields a return to education in the 5-10 percent range, a finding that has been reproduced in thousands for data sets for different countries and time periods (e.g., Psacharopoulos, 1994).

This first generation of studies on the return to education were correlational, however, and many researchers speculated the ordinary least squares (OLS) estimates of the return to education were likely biased up because of a positive connection between education and ability.[1] Starting in the early 1990s, a second generation of studies summarized in Card (1999) and Ashenfelter, Harmon, and Oosterbeck (1999) has clearly established that education had a causal effect on earnings, and that this effect was at least as large as OLS estimates obtained in first generation studies.[2] This suggest that the ability bias in OLS studies is either small, or is being offset by other factors such as measurement error in years of education.

But despite this overwhelming evidence on the benefits of education, there are some ongoing concerns in the popular press and policy circles that university education is providing training that is ill-suited for current labour market needs, and yields small and declining financial returns. For instance, in a recent budget document, the Department of Finance argues there is a growing skill gap in Canada (Finance Canada, 2014) based on an analysis of vacancy rates in occupation related to natural and applied sciences (e.g. engineers, architects), as well as skilled trades. A recent CIBC report (Tal and Enenajor, 2013) presents evidence that returns to education have been declining in Canada recently, that some university diplomas (e.g. humanities) yield small returns, and that a large fraction of university graduates in Canada earn

[1] See Griliches (1977) for an early discussion of the ability bias problem.

[2] Starting with Angrist and Krueger (1991), most of the second generation studies, including Lemieux and Card (2001) and Oreopoulos (2006) for Canada, use instrumental variable (IV) strategies as a way of estimating the causal effect of education. Related studies also indicate that education has a causal effect on a variety of outcomes variables including health (Lleras-Muney, 2005), criminal behaviour (Lochner and Moretti, 2004), and civic behaviour (Milligan, Moretti, and Oreopoulos, 2004).

less than half of the median income. More generally a common theme in the popular press is that young university graduates have a hard time finding jobs, and when they do so, these jobs often pay poorly and are unrelated to the university training (e.g. Arts graduates working as baristas).

One possible explanation for this apparent disconnect between public perceptions and the findings of the academic literature on the causal effect of education is that little remains known about the channels, or mechanisms, behind the large returns documented in the literature. A related issue is heterogeneity in the returns to education. While estimated returns are large, on average, they can be very different for different individuals. Studies of the causal effect of education on earnings provide highly credible estimates of the average effect of education on earnings, but this leaves many important questions about channels and heterogeneity unanswered.[3]

The main goal of this paper is to quantity the importance of several channels behind the effect of education of earnings, and show how they help account for large differences in the returns to education for different workers. To do so, I first consider an education-job matching framework where education has both a positive effect on productivity in a given job, and helps individuals find higher-paying jobs that highly value the skills acquired in school. This framework is closely related to "assignment" models of the labour market where more skilled workers are matched to more complex jobs that are more sensitive (in a productivity sense) to workers' skills (see, e.g. Sattinger, 1975, Teulings, 1995, and Acemoglu and Autor, 2011).

Fields of study play a central role in the matching of workers to jobs. Since workers acquire a very different set of skills in different fields of study, they should also be assigned to different type of jobs. For instance, Altonji et al. (2012) show that graduates in different fields of study tend to be clustered in very specific occupations. Given the large documented differences in earnings by field of study (e.g. Arcidiacono, 2004) and occupation (e.g. Gibbons et al., 2005), the outcome of the matching process could go a long way towards explaining why returns to education are different for different individuals. Furthermore, a number of recent studies such as Robst (2007), Nordin et al. (2010), and Yuen (2010) show that workers tend to get a large wage

[3] Imbens and Angrist (1994) show that under mild conditions (monotonicity) instrumental variables (IV) estimates, like those used in most studies of the causal effect of education on earnings, yield a consistent estimate of the local average treatment effect (for individuals marginally induced to acquire more education). In that setting, returns to education may vary across individuals depending on their ability, family background, and other factors (see Card, 1999, and Heckman and Vytlacil, 1998). But although the IV approach allows for some heterogeneity in returns, it does not help answer the question of why returns are different for different individuals.

premium when they find a job in an occupation that is a good match for the skills they acquired in their field of study.[4]

This suggests three separate reasons why education may increase earnings. First, general skills acquired as part of a university program in a given field of study increases productivity, and then wages, regardless of occupation. Second, university education helps workers get jobs in higher-paying occupations. Third, skills acquired in a given field of study are more valuable in jobs that are a good match for the education program. The first channel is a general productivity effect that occurs regardless of occupation, while the two other channels depend on the matching between education skills and occupation.

Combining data from the 2005 National Graduates Survey (NGS) and the 2006 Census, I look at the importance of the three above factors in the returns to education. Both data sources provide information about occupation and field of study. The main advantage of the NGS is that it also provides detailed information on how related one's job is to his or her education. I use this information to measure the extent of "relatedness" between each possible combination of occupation and field of study available in the data. The main disadvantage of the NGS is that it does not include individuals without post-secondary education, which precludes using this data source to look at the wage gap between university and high-school educated workers. Information about relatedness from the NGS is thus combined to the Census which contains better quality information about earnings for a much larger and representative sample of the population.

Using this combined data set approach, I first estimate the effect of the three above factors on earnings. I then perform a decomposition to quantify the contribution of the three factors (productivity and match effects) to the observed wage gap between university and high-school educated workers. The results in indicate that the productivity effect account for about half of the wage gap, while the two match effects each account for about a quarter of the gap. I also find that the all three factors play an important role accounting for the heterogeneity in the returns to education observed in the data.

The paper proceeds as follows. In Section 2, I review the existing evidence on returns to education in Canada, and conclude that those are large and relatively stable over time. In Section 3, I sketch a simple matching model and propose an empirical framework to decompose returns

[4] See also Plante (2010) who looks at this issue in the case of immigrants to Canada.

to education into a match and a productivity effect. The empirical analysis then proceeds in two steps. In Section 4, I use data from the NGS to look at the match between occupations and fields of study using a self-reported measure of how closely related one's job is to his or her field of study. I then use data from the 2006 Census in Section 5 to carry the main decomposition, and conclude in Section 6.

2. BACKGROUND: RETURNS TO EDUCATION IN CANADA

In a recent and comprehensive study based on Census data, Boudarbat et al. (2010) show that returns to education have been generally increasing in Canada since the early 1980s. For instance, they find that the university-high school wage gap has increased from 32 to 40 percentage points for men between 1981 and 2006.[5] They also show that returns to education are systematically higher for women than men, though the 1981-2006 increase in the university-high school gap is slightly smaller for women (from 45 to 51 percentage points). Boudarbat et al. (2010) also find that, consistent with Card and Lemieux (2001), the growth in the university-high school gap is substantially larger for younger than older workers. For instance, they show that the university-high school gap for men with 2 to 6 years of potential experience increased by more than 50 percent (from 31 to 48 percentage points) between 1981 to 2006, which is almost as large as the growth in the gap in the United States (Card and Lemieux, 2001).

The main advantage of the census data used by Boudarbat et al. (2010) is that both education and earnings are measured in a fairly consistent fashion over time.[6] Unfortunately, the long form census was replaced by the non-compulsory National Household Survey (NHS) in 2011, and it remains to be seen whether returns to education as measured in the NHS will be comparable to those from the census.[7] As a consequence, other data sources have to be used to look at the more recent evolution in returns to education in Canada.

[5] The university – high school gap is obtained by running a regression of log weekly wage and salary earnings of full-time workers on a quartic in potential experience and a set of dummy variables for educational achievement. The university – high school gap is the regression-adjusted gap between workers with exactly a bachelor's degree and those with exactly a high school diploma.

[6] Using other data sources such as the Survey of Consumer Finances (SCF) and the Survey of Labour and Income Dynamics (SLID), Burbidge et al. (2002) reach the conclusion that returns to education have remained stable over time in Canada. One drawback of these data sources compared to the Census data used by Boudarbat et al. (2010) is that the education question in the SCF was substantially changed in the late 1980s. Furthermore, the SCF was discontinued in 1997, which means two different data sets have to be used to cover the whole 1981-2006 period. See Boudarbat et al. (2010) for more discussion of these data issues.

[7] The main problem is that response rates in the NHS were much lower (around 70 percent) than in the long form Census (Statistics Canada, 2013).

Two recent papers by Fortin and Lemieux (2014) and Frenette and Morrissette (2014) use wage data from the Labour Force Survey (LFS) to look at recent changes in the returns to education and other dimensions of wage inequality in Canada. Both studies conclude that the returns to education have declined slightly over the last 5-10 years. For example, Fortin and Lemieux show that the (age-adjusted) university-high school gap for men and women combined declined by two percentage points between the 1998-2002 and 2008-2013 periods. Both studies also conclude that the resource boom in Western Canada was a key factor behind this reversal in previous trends in the university-high school gap. The resource boom had a positive impact on the wages of all workers, but the impact was slighter larger for high-school than university-educated workers.

These large differences in earnings between university and high-school educated workers all come from cross-sectional data sets. A recent study by Frenette (2014) shows that, as expected, there are also large differences in earnings over the life-cycle between these two groups of workers. Using a newly available data set where information about education from the 1991 census was matched to administrative income tax data, Frenette computes total cumulative earnings over the 1991-2010 period for various education groups. He finds that men with a university bachelor's degree earned on average $1,707,000 (in 2010 dollars) over this period, compared to $975,000 for high school graduates. The gap for women is also very large ($973,000 and $524,000 for university and high school graduates, respectively).

Overall, there is clear evidence that the return to education in Canada is very large. Furthermore, as discussed in the introduction, numerous studies for Canada and other countries strongly suggests that the causal effect of education on earnings is equally large. Of course, the fact that returns are large on average may hide important differences across individuals depending on their field of study, occupation, etc. These issues are systematically explored in the remainder of the paper.

3. CONCEPTUAL FRAMEWORK AND EMPIRICAL MODEL

Conceptual framework

A well-established empirical regularity is that more educated workers tend to work in occupations that pay more than typical occupations of less-educated workers. A natural explanation for this phenomenon is that more educated workers have a comparative advantage in more complex jobs in which output is more sensitive to skill (Sattinger, 1975, Teulings, 1995). This leads to a hedonic equilibrium (Rosen 1974) of the labour market where the observed relationship between education and earnings reflects both the fact that education has a positive effect on output within a given job (or occupation), and that more educated workers get assigned to more complex, and higher-paying, occupations.

In equilibrium, it is difficult to separate the role of occupational upgrading (more educated workers assigned to more complex jobs) from the "pure" return to the education (education increases productivity within a given job) since all workers of a given skill level should be assigned to the same occupation. In practice, however, getting workers of a given skill level to a given type of occupation involves a fairly complex matching process. In the presence of search frictions, not all workers will get their ideal match and we should observe a whole distribution of occupations for workers with a given skill or education level.

The matching process gets even more complex in the presence of fields of study that introduce substantial heterogeneity in type of human capital acquired through schooling. Setting up an assignment model of the labour market with heterogeneous human capital and search frictions is beyond the scope of this paper. The above discussion provides, nonetheless, a rationale for looking at the contribution of fields of study and occupational upgrading in the overall return to education.

Empirical approach

To fix ideas, consider a special case where each individual i either stops schooling after graduating from high school, or goes to university to get a bachelor's degree in a field of study (or major) j. Each field of study provides a mix of general skills that are valuable in all occupations, and some specific skills that are only valuable in occupations closely connected to the field of study. For example, specific skills acquired in a BSc in computer science are particularly valuable when the individual works as a computer programmer, but the degree also makes the individual more productive in unrelated occupations such as management.

Let the dichotomous variable $m(j,k)$ indicate whether the specific skills acquired in field of study j are valuable in occupation k, i.e. whether the occupation is a "good match" for one's education. I also refer to $m(j,k)$ as a measure of "relatedness" between occupation and education throughout the paper. This yields the following wage equation:

$$W_{ijk} = X_i\beta + b_j + c_k + \gamma \cdot m(j,k) + \varepsilon_{ijk}, \tag{1}$$

where W_{ijk} is the log wage, X_i is a standard set of covariates (e.g. gender and years of labour market experience), and ε_{ijk} is an idiosyncratic error term.

In this setting, there are three reasons why individuals who get a bachelor's degree in field of study j may earn more than those who stopped after high school. First, general skills acquired as part of a university program in field j increase wages by b_j regardless of occupation. Second, university education may help workers get a higher-paying occupation (a higher c_k). Third, holding a job in an occupation which is a good match with the field of study yields a wage premium γ. To illustrate these channels explicitly, consider the mean wage $\overline{W}_j$ for individuals in field of study j, where the covariates X_i are being ignored to simplify the exposition:

$$\overline{W}_j = b_j + \sum\nolimits_k \theta_{jk} c_k + \gamma \overline{m}_j,$$

where θ_{jk} is the fraction of individuals in field of study j who work in occupation k, and $\overline{m}_j$ is the fraction of these individuals who work in an occupation which is a good match for their field of study. Likewise, the mean wage for individuals who only hold a high school diploma is defined as:

$$\overline{W}_0 = \sum\nolimits_k \theta_{0k} c_k.$$

Note that there is no match effect for high school students, since high school is assumed to be providing general skills that are equally valuable in all occupations. The wage gap between individuals with a bachelor's degree in field j relative to those with only a high school education is given by:

$$\overline{W}_j - \overline{W}_0 = b_j + \sum\nolimits_k (\theta_{jk} - \theta_{0k})\, c_k + \gamma \overline{m}_j. \tag{2}$$

The first term on the right hand side of equation (2), b_j, is the within-occupation wage effect linked to general skills imbedded in university education in field of study j. The second term captures the occupational upgrading linked to the field of study, while the third term represents the match effect.

The total return to university education over all fields of study is obtained by averaging the field-specific returns in equation (2):

$$\overline{W}_U - \overline{W}_0 = \sum\nolimits_j \omega_j \left[b_j + \sum\nolimits_k (\theta_{jk} - \theta_{0k})\, c_k + \gamma \overline{m}_j\right], \tag{3}$$

where ω_j is the share of university-educated individuals with a diploma in field of study j.

One important empirical challenge when performing the decompositions is to consistently estimate the parameters in the main earnings regression (1). In this paper, I simply estimate the model using OLS. One may wonder, however, whether OLS estimates are biased because field of study and the match variable $m(j,k)$ are correlated with omitted factors like ability.[8]

As mentioned earlier, studies that use IV strategies to estimate the causal effect of education on earnings have generally concluded that the ability bias was small. Unfortunately, no existing studies have attempted to estimate the effect of field of study (or relatedness) using instrumental variables methods. The main challenge is that it is much more difficult to find instrumental variables for field of study and relatedness than for overall years of education that may depend on general policy variables such as compulsory schooling laws (as in Angrist and Krueger, 1991). As a result, the literature on the effect of field of study on earnings has instead tried to control for a rich set of factors such as high school grades, test scores, and other proxies for math and verbal ability. The results summarized by Altonji et al. (2012) suggests that differences in returns across fields of study tend to decline, but typically remain large and significant after controlling for these factors.

Note also that for the decomposition of the mean in equation (3), the field of study effects b_j enter together as a weighted average $\Sigma_j b_j$ that can be interpreted as the average effect of schooling on earnings, controlling for occupations and relatedness. Studies of the causal effect of education on earnings suggest that the ability bias in OLS estimates of this average effect is small. This suggests that the $\Sigma_j b_j$ component of the decomposition is not substantially biased despite the fact it is estimated using OLS.

Likewise, similar explanatory variables have been included in some of the studies on the effect of relatedness to control for the ability bias. For example, Nordin et al. (2010) include a cognitive test score based on military enlistment data (for Sweden) and find it has virtually no impact on the estimated effect of relatedness on earnings. All in all, the existing literature

[8] A related problem is that just like education in general, the choice of field of study may be signaling valuable information about ability. For instance, completing a difficult major like physics may signal a high level of general ability that may be valuable to most employers. Arcidiacono et al. (2010) suggest signaling is likely not a concern for university graduates. They find that most of the ability has already been “revealed” to the market by the time individuals complete their educational programs.

suggests that the decomposition in equation (3) is valid despite the fact the earnings equation is estimated using OLS.

4. EMPIRICAL EVIDENCE ON RELATEDNESS: NATIONAL GRADUATE SURVEY

The empirical analysis relies on two complementary data sets. The first data set is Statistics Canada's 2007 follow up of the 2005 National Graduates Survey (NGS). Every five years, Statistics Canada collects detailed information on recent college and university graduates. The most recent NGS currently available is the 2005 NGS.[9] Graduates were then followed up in 2007, at which time they were asked a battery of questions about labour market activities, educational background, and other socio-economic variables.

Unlike most other data sets like the census, the NGS also asks individuals about how related their job is to their educational qualifications.[10] This "relatedness" variable is used to create a proxy for the match variable $m(j,k)$ introduced in the previous section. The key variable used is based on the question "*How closely is the (main) job you held last week related to your certificate, diploma or degree?*" Possible answers to this question are "*closely related*", "*somewhat related*", and "*not related at all*". Robst (2007) and Yuen (2010) use similar questions from the U.S. National Survey of College Graduates and the SLID, respectively, to look at the effect of relatedness on earnings. One small difference is Robst and Yuen focus on individual-level measures of relatedness, while I focus on average measures at the occupation-field of study level (see below). Nordin et al. (2010) also use a measure of relatedness defined at the occupation-field of study level. In their case, instead of looking at self-reported measures of relatedness, their construct a measure of mismatch based on a comparison between occupational and field of study classifications. Interestingly, both approaches yield similar results.

The NGS also asks whether individuals feel they are overqualified for their jobs. Although this variable is not as directly related to the question of matching (between occupation and field of study), I will later show that it is strongly correlated with the relatedness variable.

In the public use version of the NGS, both field of study and occupation are reported at a highly aggregated level. There are ten categories available for field of study, and nine for occupations. This yields 90 possible of field of study – occupation categories. While it could be

[9] Microdata from the 2010 NGS was not yet available at the time this paper was written.
[10] See Boudarbat and Chernoff (2012) for an empirical analysis of job relatedness based on the 2005 follow up to the 2000 NGS, and Yuen (2010) who uses the Survey of Labour and Income Dynamics.

useful to have more detailed categories for field of study and occupation, this must be balanced against the fact that there are 10,925 usable observations in the NGS. This means there are slightly more than 100 observations, on average, for each of the 90 field of study – occupation categories available in the public use files. Using a more detailed breakdown would be challenging as the number of observations in many fields of study – occupation categories would become quite small.

One possible drawback of using this highly aggregated approach is that individuals that look well matched to their job at the aggregate level may not be well matched at the more detailed level. As a result, some of variation in the relatedness variable within aggregate field of study – occupation groups may be due to the fact that controls for field of study are not detailed enough, which could yield biased estimates. I show below that this does not appear to be a problem since this estimated effect of relatedness is robust to alternative estimation strategies based on between- and within-group variation in relatedness.

The public use files of the NGS also provide a measure of annual earnings (in $5,000 bins) for the job held during the survey week in 2007.[11] I use this earnings measure in some of the analysis reported below, though the main results of the paper rely on census data instead. In addition to this limited measure of earnings, a major drawback of the NGS is that it does not include any information about individuals who do not have a post-secondary diploma.

Table 1 shows a number of descriptive statistics from the NGS. As is well known, women are over represented among graduates from post-secondary institutions. Table 1 (last column) shows that over 60 percent of graduates in 2005 are women. The more detailed breakdown of field of study by gender indicates some stark differences between men and women. Close to a quarter a men studied in engineering and related fields (technology and architecture), compared to only three percent for women. Men are also largely over-represented in computer science and mathematics (9 percent compared to 2 percent for women). By contrast, 20 percent of women are in health, compared to only 7 percent for men. The fact that

[11] There are 18 bins $5,000 bins (from less than $5,000 to $85,000 and more) available. A proxy for annual earnings is constructed using mid-points of the $5,000 intervals for bins up to $85,000. For earnings above $85,000, earnings are imputed under the assumption that the top end of the distribution follows a Pareto distribution with a parameter α=3.5, as in Lemieux 2006. This implies that average earnings above the topcode are equal to 1.4 times the topcode. This yields an imputed earnings value of 1.4 x $85,000 = $119,000 for the top earnings category ($85,000 and more).

engineering and related fields is the highest paying field of study (last row of Table 1) suggests that gender differences in field of study explains some of the gender wage gap.[12]

Table 1 also indicates some important differences in field of study depending on the level of schooling. For instance, health and engineering account for a larger share of diplomas at the college (community college or CEGEP in Quebec) than the university level. By contrast, education, humanities, and social sciences and law are much more popular at the university level.

Panel B of Table 1 shows information on what individuals think of the connection between their job and education qualifications. The first column indicates that 61 percent of workers think their job is strongly related to their education. This jumps to more than 80 percent when the definition of relatedness also includes those who think their job is somewhat related to their education. There are also some large differences in relatedness by field of study. More than 90 percent of workers who studied in health and education say their job is strongly or somehow related to their education. Computer science, mathematics, and engineering are close behind at around 90 percent. By contrast, only 50 percent of individuals with a diploma in the humanities feel that their job is closely or somehow related to their education.

This relatedness measure is also closely connected to the measure of overqualification reported in the third row of Panel B. Less than 20 percent of workers with a diploma in education or health feel they are overqualified for their job, compared to close to 50 percent in the humanities.[13]

One obvious reason why a job may not be related to one's education is that it does not utilise the specific skills acquired in an educational program (e.g. a history graduate working as a barista). Table 2 shows the distribution of occupations for each of the ten fields of study. The table reports evidence for all levels of schooling pooled together, but qualitatively similar results are obtained when looking at the sub-sample of individuals with a bachelor's degree in Appendix Table 1. For some fields of study, most graduates are concentrated in occupations closely related

[12] This is confirmed by running simple regressions of (log) earnings on gender with and without field of study. The raw gender gap in log earnings (only adjusting for age and education in a regression) is 0.205, and goes down to 0.165 after controlling for differences in field of study.

[13] While overqualification could be interpreted as another dimension of the mismatch between skills and occupations, in this paper I focus instead on the relatedness issue which is more closely connected to the match between field of study and occupations.

to their educational programs.[14] For example, 85 percent of education graduates work in education (or social sciences which are part of the same one-digit occupation category). Likewise, three quarters of workers with a diploma in health work in health occupations, while a large fraction of math, computer science, and engineering graduates work in broadly defined "science" occupations.

More generally, since occupations and field of study are ranked in relatively comparable order, we should expect most observations to lie around the main diagonal of the table if there was a good match between field of study and occupations.[15] To informally see whether this is supported in the data, Table 2 highlights in bold characters the occupation-field of study cells for which the proportion of workers in the occupation is more than twice as high as in the marginal distribution (for all fields of study combined). Most of the highlighted numbers lie close to the main diagonal, suggesting a systematic connection between field of study and occupations.

The connection between occupations and field of study is explored more systematically in the last column of Table 2. The column reports the concentration of workers in different occupations as summarized by the Duncan index of occupational segregation. The Duncan index indicates the fraction of workers in a given field of study who would have to change occupations to get the marginal distribution of occupations for the whole workforce. For instance, 72.7 percent of workers with health diplomas work in health occupations, compared to only 12.3 percent for the whole workforce (last row of Table 2). This means that we would need to move 72.7 – 12.3 = 60.4 percent of workers out of health occupations, and redistribute them to the other occupations to get the marginal distribution of occupations at the bottom of the table. More generally, the Duncan index DI_j in field of study j is defined as:

$$DI_j = \frac{\sum_k |\theta_{jk} - \theta_k|}{2},$$

where θ_{jk} is the fraction of individuals in field of study j who work in occupation k and θ_k is the fraction of all individuals who work in occupation k (the marginal distribution).

Consistent with the evidence discussed above, there are substantial differences in the Duncan index for different fields of study. It ranges from around 60 percent in engineering,

[14] Altonji et al. (2012) also show that graduates in different fields of study tend to be clustered in very specific occupations. Nordin et al. (2010) use this feature of the data to construct their measure of relatedness by manually deciding whether one's job is closely related to his or her field of study.
[15] As can be seen on Table 2, there is a close but not one-to-one correspondence between occupations and field of study, since these two variables are based on quite different classification systems.

health, and education, to only about 20 percent in humanities and physical and life sciences (mostly biology). Differences in the Duncan index across field of study mirror those in the measure of relatedness reported in Table 1. The correlation between these two statistics is large and positive (0.76). This suggests that the main reason many graduates in fields like the humanities have jobs not related to their education is that they have a harder time finding jobs in occupations where their specific skills are highly valued.

This is shown explicitly in Table 3, which reports the average relatedness (fraction of workers reporting their job is strongly or somehow related to their education) for each combination of occupation and field of study.[16] As expected, the results show that the measure of relatedness is particularly high when workers have jobs in occupations closely connected to their field of study. For example, 99 percent of workers with a diploma in health who work in health report that their job is closely or somehow related to their education. This fraction drops to less than 50 percent if they work in sales occupations instead. Interestingly, the same pattern can be observed for fields of study like arts and communications for which graduates report a lower level of relatedness over all occupations (last column of Table 3). When these graduates work in "arts" occupations, their level of relatedness is very high at 95 percent, but it drops to 35 percent when they work on sales occupations instead.

The evidence reported in Tables 2 and 3 suggests that the relatedness variable is a good measure of mismatch between jobs and skills. Measured relatedness tends to be high when workers have a job that is a good match for their skills (e.g. engineers working in engineering), and low when they work in jobs much less related to their skills (e.g. sales jobs). This provides a strong empirical foundation for using relatedness as the key measure of mismatch in the decomposition exercise presented in the next section.

One final point about relatedness is that, despite oft heard claims that more technical diplomas and trade certificates provide training better geared to the labor market than university education, Figure 1 shows there is no systematic differences in relatedness between college and university (bachelor's degree) graduates. Likewise, university graduates do not report being more overqualified for their jobs than college graduates. In fact, workers who have completed a

[16] As in the case of Table 2, Table 3 shows results for all levels of schooling pooled together. Similar results are obtained using the measure of "strong relatedness" (instead of strongly or somehow related in Table 3) in Appendix Table 2, or the sub-sample of individuals with a bachelor's degree in Appendix Table 3.

masters, doctorate, or professional degree (graduate school category in the figure) are the most likely to have a job strongly related to their education.

Relatedness and Earnings

While the NGS cannot be used to carry out a decomposition of the university – high school gap, it can be used to provide some evidence on the effect of relatedness, occupations, and field of study on earnings.

Table 4 reports average log earnings for the same occupation-field of study cells as in Tables 2 and 3. To ease exposition, average earnings by occupation-field of study cells are reported as deviations relative to average earnings for all workers combined. Cells for which average relatedness (fraction of workers reporting their job is strongly or somehow related to their education) is larger or equal to 95 percent are highlighted in bold characters.

Table 4 shows a strong relationship between relatedness and earnings. Most of the high-relatedness cells highlighted in bold tend to have substantially above average earnings. For instance, engineering graduates working in science occupations (relatedness of 96 percent in Table 3) earn 27 percent above average. More importantly, this group of individuals earns more than engineering graduates in all occupations (19 percent above average) and workers from all fields of study working in science occupations (22 percent above average). This suggests a positive earnings premium for relatedness that goes above and beyond the fact that workers in high paying occupations or fields of study tend to report a higher level of relatedness.

This last point is explored more systematically in Table 5 which presents a set of earnings regression that all control for gender, schooling level, and field of study. Since individuals are followed up two years after graduation, the earnings measure provides only an early snapshot into one's career. There is, nonetheless, clear evidence at this early point that schooling has a large and positive effect on earnings, and that women earn substantially less than men.

Column 1 also shows very large differences in earnings across fields of study. For example, engineering graduates earn 34 percent more than social science and law graduates (the base group in the regression). This particular earnings gap is larger than the 29 percent gap between workers with a bachelor's degree and a college degree. This is consistent with the literature that systematically shows very large differences across fields of study.[17]

The set of nine occupation dummies are added to the regression model in column 2. Adding occupations to the model reduces the effect of field of study, which is not surprising given the correlation between occupations and field of study (Table 2). For instance, the engineering premium declines from 34 to 19 percent once occupations are included in the model. More generally, the standard deviation of the field of study effect decreases from 0.172 to 0.144 after including occupation dummies in the model.

The relatedness variable is added to the regression model in column 3. The measure of relatedness used in columns 3-5 is a dummy variable indicating whether an individual's job is either strongly or somehow related to his or her education. Consistent with the descriptive evidence reported in Table 4, the effect of relatedness is large (over 30 percent) and significant. As discussed at the end of Section 3, one may wonder whether the effect of relatedness is biased up because it is correlated with omitted factors like ability. Perhaps "good", or more able, workers have an easier time finding a job that uses the specialised skills they learn in school, and that these workers would have earned more than their peers regardless of the job they found.[18] To address at least part of this bias, I include in column 4 two variables that are arguably related to ability. The first variable is a self-reported ranking variable based on the question "Compared to the rest of your graduating class in your field(s) of study, did you rank academically in the top 10%, 10-25%, 25-50%, or below the top half?". While this variable should be a good proxy for academic ability, Appendix Figure 1 shows that students largely overestimate their position in the distribution.[19] The other control added in column 4 is a dummy variable indicating whether the individual was part of a co-op program, which should in principle help good students find a job in their field. The results reported in column 4 indicate that including the rank and co-op

[17] See for instance Arcidiacono (2004), Moussaly-Sergieh and Vaillancourt (2009), and Altonji et al. (2012). One small difference compared to most existing studies is that I pool all post-secondary graduates here, which means that the effect of field of study is a mix of effects for the three education groups (college graduates, undergraduates, and individuals who went to graduate school). Results are similar but less precise when I limit the analysis to undergraduates.
[18] See also Montmarquette et al. (2002) and Stinebrickner and Stinebrickner (2014) for more evidence on how individuals with different abilities and backgrounds choose their field of study.
[19] Since the NGS sample is a random sample of graduates, we should more or else observe 10 percent of individuals in the top 10 percent of their classes, but close to 40 percent of them claim there are in this group. Likewise only 2 percent of individuals report that they are in the lower half of the distribution, a far cry from the 50 percent of individuals who should be there in a random sample.

variables in the regression has virtually no impact on the estimated effect of relatedness. This is consistent with Nordin et al. (2007) who also find that the estimated effect of relatedness remains the same when controls for ability are added to the earnings regression.

Note also that since the regression models also include occupation and field of study controls, the effect of relatedness is not driven by the fact individuals in some higher paying fields (like engineering) are more likely to have a job related to their education. Conditional on field of study and occupation, there are two reasons why relatedness may have a positive effect on earnings. First, within a given occupation-field of study cell, individuals who report having a job related of their education may be earning more than individuals who say their job is not related to education. Column 5 reports this “within” effect by adding a full set of occupation-field of study interactions to the regression. The results indicate that doing so has no impact on the estimated effect of relatedness.

The second reason has to do instead with between occupation-field of study differences in relatedness. These differences were illustrated earlier in Table 3. The “between” estimate is obtained by replacing individual measures of relatedness with cell means (average relatedness in the occupation-field of study cell). In that case the resulting estimated effect of relatedness solely relies on the fact some occupation-field of study combinations are good matches, while others are not. The estimate is identified from the interaction between occupation and field of study dummies, since main effects (for occupations and field of study) are also included in the regression.

The between estimate is arguably more robust to the ability bias problem than the within estimate since it only relies on variation at the group level. Another advantage over the between estimate is that it is much clearer where the variation is coming from. In the case of the within estimator, one may wonder why it is that some individuals who work in the right occupation given their education end up reporting having a job not related to their education. By contrast, in the case of the between estimator it is clear why some occupation–field of study cells exhibit a high level of relatedness (e.g. educators working in education), while others do not (e.g. educators working in retail sales).

The between estimator is implemented in practice using an IV procedure where the full set of interactions between occupations and field of study are used as instruments for relatedness.[20] The estimates reported in column 6 of Table 5 yield a slightly larger estimated effect of relatedness (35 percent). Not surprisingly, standard errors increase when only the between variation is used. The effect of relatedness remains statistically significant, nonetheless, though the difference relative to the OLS estimate is not statistically significant (p-value of the Hausman test is 0.604 when comparing the models in columns 3 and 6). Note that standard errors are clustered at the occupation-field of study level since this is the only source of variation in relatedness used in the between model. For the sake of comparability, standard errors are also clustered at the occupation-field of study level in the other models reported in Table 5.

It is reassuring to see that all the estimated effects of relatedness reported in Table 5 are similar in magnitude since some of them, such as the within estimate in column 5, are arguably more affected by the ability bias problem. This mirrors the findings of earlier studies that have used both the within (Robst, 2007, and Yuen, 2010) and between (Nordin et al, 2010) approach to estimate the effect of relatedness on earnings, and found similar effects. The next step of the analysis consists of merging the measures of relatedness from the NGS (at the occupation-field of study level) to the Census data to compute the decomposition introduced in Section 3.

5. DECOMPOSITION OF THE GAP BETWEEN UNIVERSITY AND HIGH SCHOOL EDUCATED WORKERS

2006 census data

The main earnings analysis is based on the public use file of the 2006 Census. Since 1986, the census has been collecting information on field of study for individuals who hold a post-secondary diploma. Both the 2006 census and the NGS use the same classification of instructional programs (CIP). The main results reported below are based on an analysis of earnings from the census merged with information on the quality of the field of study – occupation match (i.e. relatedness) obtained from the NGS. Since information on relatedness is obtained by matching data at the field of study – occupation level, the estimated effect of relatedness on earnings only relies on the between-group variation discussed in the previous section. This yields less precise estimates since within-group variation is not available in this

[20] Including cells means directly in the regression yields an estimated effect of relatedness that is almost identical to the IV estimate (0.349 vs. 0.346).

case. But as I show below precision is not a major issue since, as in the case of the NGS, the estimated effect of relatedness in the Census is large and statistically significant.

The "long form" of 2006 census provides detailed information on earnings, weeks of work, and part-time / full-time status during the previous year (2005). In the 2006 census, respondents were given the option, for the first time, to share their income tax information with Statistics Canada instead of self-responding to the income questions in the Census. Over 80 percent of respondents agreed to do so. As a result, the quality of the information about earnings is excellent, as it is mostly based on administrative income tax data.

It is not possible to construct a measure of hourly wages in the Census because detailed information on annual hours of work is not available. Following the existing literature (e.g. Boudarbat et al., 2010) I use the weekly earnings (wages and salaries plus self-employment earnings) of full-time workers as the principal measure of wages. I also present results using only wage and salary earnings. I trim all wage observations with weekly earnings below $100 because they imply implausible values for hourly wages.[21] Although earnings are top-coded in all public use files of the Census, no adjustment for top-coding is required since Statistics Canada used a procedure that yields valid estimates of average earnings for top earners in the 2006 Census.[22]

The decomposition exercise presented below focuses on a comparison of individuals whose highest diploma is a bachelor's degree to those with a high school diploma (and no further diploma). I will refer to these two groups as "university" and "high school" from now on. To allow for enough time to complete a bachelor's degree, the university sample consists of individuals aged 25 to 64 at the time of the Census (June). In order to compare individuals with about the same level of potential experience --in the tradition of Mincer (1974)-- I restrict the high school sample to individuals age 20 to 59, but also report additional results where the age

[21] Since full-time workers are defined as working at least 30 hours a week, a full-time worker earning $100 a week makes at most $3.33 an hour. This represents less than half of the minimum wage in most provinces in 2005, the exceptions being Newfoundland and Labrador (minimum wage of $6.25 at the end of 2005) and New Brunswick (minimum wage of $6.30).

[22] Prior to the 2006 Census, earnings in the public use files were censored at an arbitrary level (e.g. $200,000 in 2001) which lead to an understatement of earnings at the top in unadjusted earnings data. In the 2006 public use files of the census, Statistics Canada used instead the average earnings among the top one percent of earners (by census metropolitan area and gender) to impute earnings for these top earners. The procedure insures that average earnings for all workers are no longer affected by the top-coding procedure. Note, however, that there may still be a small bias in estimates of the wage gap by education groups, since the procedure used by Statistics Canada implicitly assumes that individuals in the top one percent all have the same average earnings regardless of education achievement. Lemieux and Riddell (2014) show that this assumption is not supported in data based on the Census master files.

cutoffs for high school graduates are the same as university graduates (age 25-64).[23] I also exclude immigrants from the sample, since their education is often not comparable to the education of Canadian-born workers.

Descriptive statistics and earnings regressions

Figure 2 shows the distribution of earnings for high school and university graduates. The figure is computed by first finding the earnings quartile cutoffs for workers of all education levels. I then look at the fraction of high school and university graduates who fall into these four quartiles. The figure shows that the earnings distribution of university graduates is substantially above the one for high schools graduates. For instance, only about 10 percent of university graduates are in the lowest earnings quartile, compared to about a third for university graduates. Likewise, about 15 percent of high school graduates are in the top quartile, compared to over 40 percent for university graduates.

Within university graduates, there are also large differences in the earnings distribution across fields of study, as in the NGS data. Figure 3 compares the distribution for the highest (engineering) and lowest (fine arts and communication) paying fields of study relative to high school graduates. The figure shows that arts graduates earn barely more than high school graduates, and much less than engineering graduates. For example, 65 percent of engineering graduates are in the top earnings quartile, which is four time as large as high school graduates (17 percent), but also three times as large as arts graduates (22 percent).

These large differences across field of study can also be seen in the regression models reported in Table 6. Column 1 shows the estimates from a model with field of study dummies, as well as a set of controls for potential experience and gender, while column 2 adds occupation dummies and the relatedness variable to the specification.[24] Unlike the NGS where the reference

[23] In most undergraduate programs outside Quebec it normally takes four years to graduate. In Quebec there is normally a five-year gap between high school and university graduation (two years in CEGEP and three years in university). Furthermore, some programs take an extra year and many students take more than the "normal" time to graduate. Accordingly, assuming that Canadian university graduates finish their studies five years, on average, after completing high school is a reasonable approximation. Prior to 2006 it would have been possible to construct a more accurate measure of potential experience (or age at graduation), but this is no longer possible in the 2006 census that did not ask explicitly about years of schooling. See Boudarbat et al. (2010) for more discussion of this issue.

[24] Relative to the NGS, a more detailed set of 25 occupations are available in the public use file of the 2006 census, and 24 occupations dummies are included in the regression models. The controls for potential experience are a set of dummies based on education and age categories (5-year age categories available in the public use file of the census).

group for field of study was social sciences and law, in the census regression I use high school as the reference group (field of study is only asked to individuals with some post-secondary education). Thus, the field of study effects reported in Table 6 represent the wage gap between individuals with a bachelor's degree in a given field relative to high school graduates.

Consistent with Figure 3, the results reported in column 1 of Table 6 indicate that the return to a university diploma in engineering is much larger (0.743) than in the arts (0.189). The returns decline substantially once occupation dummies and the relatedness variable are added in column 2, though the ranking across fields of study remains relatively unchanged. One important exception, however, is social sciences and law that becomes the highest paying field once occupation and relatedness are controlled for. I discuss this in more details below, since the occupation upgrading and matching effects captured by the occupation and relatedness dummies are the key mechanisms analysed in the decomposition.

It is not possible to get a direct measure of relatedness for high school graduates since they are not part of the NGS sample. Since the skills acquired in high school are fairly general in nature (writing, math, etc.), one could arbitrary assume that relatedness for high school graduates is comparable to relatedness in a university major that mostly provides general skills. A natural candidate is the humanities that have the lowest level of relatedness of all majors (51.4 percent, see Table 1) and arguably provide much more general skills (critical thinking, writing, etc.) than, say, health programs. Therefore, the measure of relatedness used in column 2 is set to 51.4 percent for high school graduates.

Column 2 shows that the effect of the relatedness variables in the Census is 0.49, which is even larger than in the NGS (as in the regression models for the NGS, standard errors are clustered at the occupation-field of study level). This is a very large effect. Since relatedness reaches close to 100 percent for some combinations of occupations and fields of study (e.g. engineers work in science jobs), differences in relatedness could account for a 25 percentage points earnings difference between these groups and typical graduates from the humanities (average relatedness of 51 percent). The connection between earnings and relatedness is also illustrated in Figure 4 which plots average earnings against average relatedness by occupation

---

Following the above discussion, the first potential experience category is the age 20-24 category for high school graduates, and the age 25-29 category for university graduates, and so on and so forth.

and field of study. There is a clear positive slope in the figure, indicating that the regression results are not just driven by a few outliers.

As an alternative to the assumption that relatedness for high school graduates is the same as for humanities graduates, I consider an alternative procedure where relatedness is summarized by a dummy variable. An occupation-field of study pair is classified as having a high level of relatedness (dummy equal to 1) if it is above mean relatedness (0.783) for the whole NGS sample, and is classified as having a low level of relatedness otherwise (dummy equal to 0). In that case the relatedness dummy is simply set to zero for high school graduates. Another advantage of just using a dummy variable is that it simplifies the decomposition exercise, and corresponds to the model presented in equation (1) where the relatedness (or matching) term $m(j,k)$ is also a dummy variable.

Column 3 shows that the estimated effect of the relatedness dummy is close to 20 percent, and that the fit of the model (R-square) is comparable to the case where the continuous measure of relatedness is used instead. The rest of Table 6 then shows a variety of alternative specifications for men and women considered separately (columns 4 and 5), young workers with less than ten years of potential experience (column 6), age instead of potential experience controls (column 7), and wage and salary workers only (column 8). While there are differences in the effect of relatedness across different specifications, the effect is systematically large and significant.

The estimated effect of field of study on earnings is also relatively stable across the various specifications reported in columns 3 to 8. The only noticeable pattern is that the estimated effects are systematically smaller when age controls are used instead of experience controls in column 7. Since the field of study effects are defined relative to high school graduates, this reflects the well-known fact (Mincer 1974) that controlling for age instead of experience underestimates the returns to education. The problems is that controlling for age fails to account for the fact that high-school graduates of a given age have more labour market experience than university graduates of the same age.

Decomposition results

Table 7 presents the results of the decomposition of the mean wage gap outlined in equation (3). The decomposition is computed using the model reported in column 3 of Table 6 that also adjusts for experience (5-year categories) and gender. The “total gaps” reported in column 2 of Table 7 range from 0.177 in arts and communications to 0.747 in engineering. The total gap for all fields combined using the weights reported in column 1 is equal to 0.554. This is larger than many other estimates reported in the literature for a number of reasons. First, both men and women are combined in the main analysis sample, and returns are systematically larger for women than men (Boudarbat et al., 2010). Second, many university-educated immigrants have relatively low earnings, and excluding them as I do here increases substantially the estimated return. Third, controlling for experience instead of age (or no controls at all) also substantially increases the estimated return to education.

The next three columns show the key elements of the decomposition. The occupation upgrading effect reported in column 3 accounts for 0.119, or 21.5 percent, of the overall university-high school gap. The occupation upgrading effect also tends to be larger in fields in which the total gap is also larger. For instance, it is equal 0.164 in engineering but only 0.009 in arts and communications. Interestingly, the largest occupational upgrading effect is in business (0.185). This reflects the fact that business graduates are more likely than others to work as senior managers, the highest paying of all occupations included in the regression models.

The match, or relatedness, effect reported in column 4 is simply the product of the relatedness effect (0.182) and the fraction of individuals working in occupations related to their field of study. Column 4 shows that fields of study can be divided in two distinct groups in terms of this match effect. The effect is quite high at around 0.160 in education, health, business, and most fields in pure and applied science. The exception in the latter case is physical and life sciences (mostly biology) for which the match effect is substantially lower at 0.095. The match effect is even lower in the arts, humanities and social sciences. Overall, the match/relatedness effect accounts for 22.3 percent of the university-high school gap.

The residual component left after accounting for the match and occupational upgrading effect is the “pure” return reported in column 5. Looking at all fields combined, the pure return accounts for slightly more than a half of the total gap. Some interesting patterns also emerge when looking at different fields separately. In particular, the total gap for the humanities (0.370) is about 25 percentage point lower than in business (0.626) or health (0.632). Most of this difference is explained, however, by occupational upgrading and match effects, and the pure return in the humanities is only 5 percentage points lower than in business and health. This suggests that the general skills provided by a degree in the humanities are as valuable as in health or business. The difference is that humanity graduates are less likely to end up in high-paying occupations, or in occupations that make good use of the specific skills they learned in university.

One key take-away point is that there is a great deal of heterogeneity in the returns to education depending on field of study, occupation, and to what extend workers' occupations are related to their fields of study. Looking at Table 7 again, graduates in arts or communication working in occupations similar to high school graduates, and not related to their own education, earn only 8.5 percent more than high school graduates (the 0.085 pure return in column 5). At the other end of the spectrum, engineering graduates with the average occupational upgrading and relatedness of their field of study earn 75 percent more (column 2) than high school graduates.

It is remarkable to observe such a range of variation in the return to education based on observable characteristics. This means it is relatively easy to find a group of university graduates who earn barely more than high school graduates, despite the fact the average return for all university graduates is much larger.

5. CONCLUSION

This paper considers several possible channels behind the well-documented effect of education on earnings. The first or "pure returns" channel is that education makes workers more productive on a given task, as in a conventional human capital framework. The second channel is based on the idea that education helps workers get assigned to higher-paying occupations where output is more sensitive to skill. A third and final channel is that workers are more productive and earn more when they are matched to a job related to their field of study. Using data from the 2006 census, I find that the two latter channels account for close to half of the conventional return to education. As shown in the online appendix, a similar conclusion about the relative importance of the two factors is reached when looking at the whole distribution of earnings instead of just mean earnings.

The results also indicate that the return to education varies greatly depending on occupation, field of study, and the match between these two factors. It ranges from less than 10 percent for arts graduates with no occupational upgrading who work in an occupation unrelated to their education, to more than 75 percent for engineering graduates with an average occupational attainment for workers in that field. This means it is relatively easy to find a group of university graduates who earn barely more than high school graduates, despite the fact the average return for all university graduates is much larger. This may explain why we often hear popular press stories about university graduates working as baristas and earning barely more than high school graduates, despite the well-known fact that, on average, the payoff to higher education is very large.

One potential pitfall of the decomposition procedure used in this paper is that both earnings and the choice of field of study and occupation may depend on unobservable factors such as ability. Little attempt is made to deal with this problem here, in part because the large literature on the causal effect of education on earnings suggests that the ability bias is small. Data permitting, it would be a worthwhile endeavour to see whether the main conclusions reached in this paper remain after controlling for a richer set of confounding factors.

REFERENCES

Acemoglu, Daron, and David H. Autor (2011) 'Skills, tasks and technologies: Implications for employment and earnings,' *Handbook of labor economics*, vol. 4, ed. O. Ashenfelter and D. Card (Amsterdam: North-Holland)

Altonji, Joseph G., Erica Blom, and Costas Meghir (2012) 'Heterogeneity in Human Capital Investments: High School Curriculum, College Major, and Careers,' *Annual Review of Economics* 4, 185-223

Angrist, Joshua D., and Alan B. Krueger (1991) 'Does Compulsory School Attendance Affect Schooling and Earnings?,' *Quarterly Journal of Economics* 106, 979-1014

Arcidiacono, Peter (2004) 'Ability sorting and the returns to college major,' *Journal of Econometrics* 212, 343-75

Arcidiacono, Peter, Patrick Bayer, and Aurel Hizmo (2010) 'Beyond Signaling and Human Capital: Education and the Revelation of Ability,' *American Economic Journal: Applied Economics* 2, 76-104

Ashenfelter, Orley, Colm Harmon, and Hessel Oosterbeek (1999) 'A review of estimates of the schooling/earnings relationship, with tests for publication bias,' *Labour economics* 6, 453-70

Boudarbat, Brahim, and Victor Chernoff (2012) 'Education--job match among recent Canadian university graduates,' *Applied Economics Letters* 19, 1923-26

Boudarbat, Brahim, Thomas Lemieux, and W. Craig Riddell (2010) 'The Evolution of the Returns to Human Capital in Canada, 1980-2005,' *Canadian Public Policy* 36, 63-89

Burbidge, John B., Lonnee Magee, and A. Leslie Robb (2002) 'The Education Premium in Canada and the United States,' *Canadian Public Policy* 28, 203-17

Card, David (1999) 'The causal effect of education on earnings,' in *Handbook of labor economics*, vol. 3, ed. O. Ashenfelter and D. Card (Amsterdam: North-Holland)

Card, David, and Thomas Lemieux (2001) 'Can Falling Supply Explain the Rising Return to College for Younger Men? A Cohort-Based Analysis,' *Quarterly Journal of Economics* 116, 705-46.

Department of Finance Canada (2014) *Jobs Report: The State of the Canadian Labour Market*

Fortin, Nicole M., and Thomas Lemieux (2014) 'Changes in Wage Inequality in Canada: An Interprovincial Perspective,' *Canadian Journal of Economics*, forthcoming.

Frenette, Marc (2014) 'An Investment of a Lifetime? The Long-term Labour Market Premiums Associated with a Postsecondary Education,' Statistics Canada Analytical Studies Branch Research Paper No. 359

Frenette, Marc, and René Morissette (2014) 'Wages and Full-time Employment Rates of Young High School Graduates and Bachelor's Degree Holders, 1997 to 2012,' Statistics Canada Analytical Studies Branch Research Paper No. 360

Gibbons, Robert, Lawrence Katz, Thomas Lemieux and Daniel Parent (2005) 'Comparative Advantage, Learning, and Sectoral Wage Determination' *Journal of Labor Economics* 23, 681-724

Griliches, Zvi (1977) 'Estimating the Returns to Schooling: Some Econometric Problems,' *Econometrica* 45, 1-22

Heckman, James, and Edward Vytlacil (1998) 'Instrumental variables methods for the correlated random coefficient model: Estimating the average rate of return to schooling when the return is correlated with schooling,' *Journal of Human Resources* 33, 974-987

Imbens, Guido, and Joshua Angrist (1994) 'Identification and estimation of local average treatment effects,' *Econometrica* 62, 467-475.

Lemieux, Thomas, and David Card (2001) 'Education, Earnings, and the Canadian G.I. Bill,' *Canadian Journal of Economics 34*, 313-44

Lemieux, Thomas, and W. Craig Riddell (2014) 'Top Incomes in Canada: Evidence from the Census,' Working paper, February.

Lleras-Muney, Adriana (2005) 'The relationship between education and adult mortality in the United States,' *The Review of Economic Studies* 72, 189-221

Lochner, Lance, and Enrico Moretti (2004) 'The Effect of Education on Crime: Evidence from Prison Inmates, Arrests, and Self-Reports,' *American Economic Review* 94, 155-189.

Milligan, Kevin, Enrico Moretti, and Philip Oreopoulos (2004) 'Does education improve citizenship? Evidence from the United States and the United Kingdom,' *Journal of Public Economics* 88, 1667-1695.

Mincer, Jacob (1974) *Schooling, Experience and Earnings* (New York: Columbia University Press)

Montmarquette, Claude, Kathy Cannings, and Sophie Mahseredjian (2002) 'How do young people choose college majors?' *Economics of Education Review* 21, 543-56

Moussaly-Sergieh, Karim, and François Vaillancourt (2009) 'Extra Earning Power: The Financial Returns to University Education in Canada,' e-briefs 79, C.D. Howe Institute

Nordin, Martin, Inga Persson, and Dan-Olof Rooth (2010) ‘Education–occupation mismatch: Is there an income penalty?’ *Economics of Education Review* 29, 1047-59

Oreopoulos, Philip (2006) ‘The Compelling Effects of Compulsory Schooling: Evidence from Canada,’ *Canadian Journal of Economics* 39, 22–52.

Plante, Johanne (2010) ‘Characteristics and Labour Market Outcomes of Internationally-educated Immigrants’ Statistics Canada, Culture, Tourism and the Centre for Education Statistics, Working Paper No. 84.

Psacharopoulos, George (1994) ‘Returns to investment in education: A global update,’ *World Development* 22, 1325-43

Robst, John (2007) ‘Education and job match: The relatedness of college major and work,’ *Economics of Education Review* 26, 397-407

Rosen, Sherwin (1974) ‘Hedonic prices and implicit markets: product differentiation in pure competition,’ *Journal of Political Economy* 82, 34-55.

Sattinger, Michael (1975) ‘Comparative advantage and the distributions of earnings and abilities,’ *Econometrica* 43, 455-468.

Statistics Canada (2013) ‘Income Composition in Canada, 2011 National Household Survey,’ Catalogue no. 99-014-X201100, Ottawa.

Stinebrickner, Ralph and Todd Stinebrickner (2014) ‘A Major in Science? Initial Beliefs and Final Outcomes for College Major and Dropout,’ *The Review of Economic Studies* 81, 426-72

Tal, Benjamin, and Emanuella Enenajor (2013) “Degrees of Success: The Payoff to Higher Education in Canada”, CIBC Economics “In Focus”

Teulings, Coen N. (1995) ‘The wage distribution in a model of the assignment of skills to jobs,’ *Journal of Political Economy* 103, 280-315.

Yuen, Jennifer (2010) ‘Job-education Match and Mismatch: Wage Differentials” *Perspective on Labour and Income* 11, 1-26.

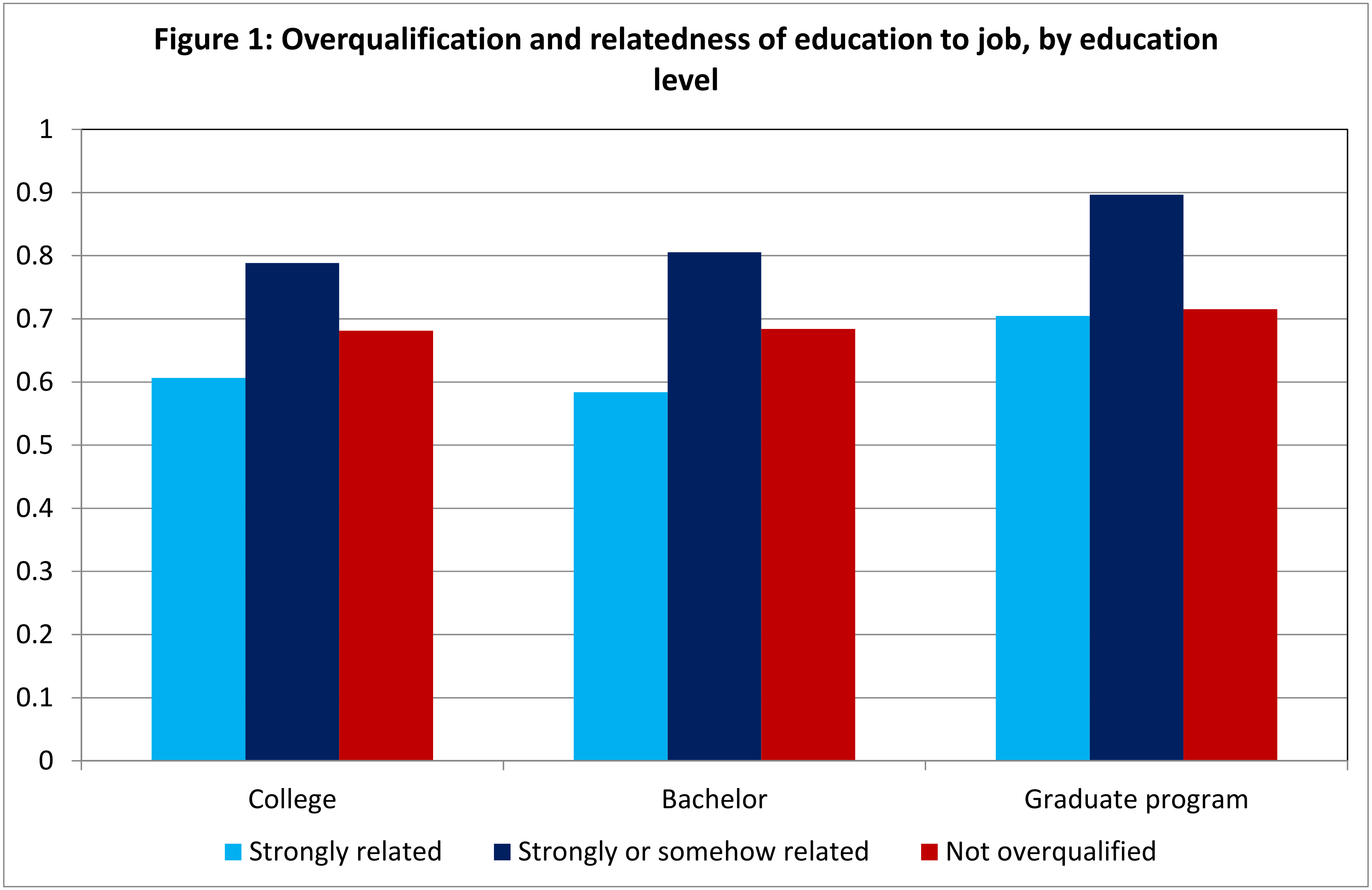
Figure 1: Overqualification and relatedness of education to job, by education level
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
College
Bachelor
Graduate program
Strongly related
Strongly or somehow related
Not overqualified

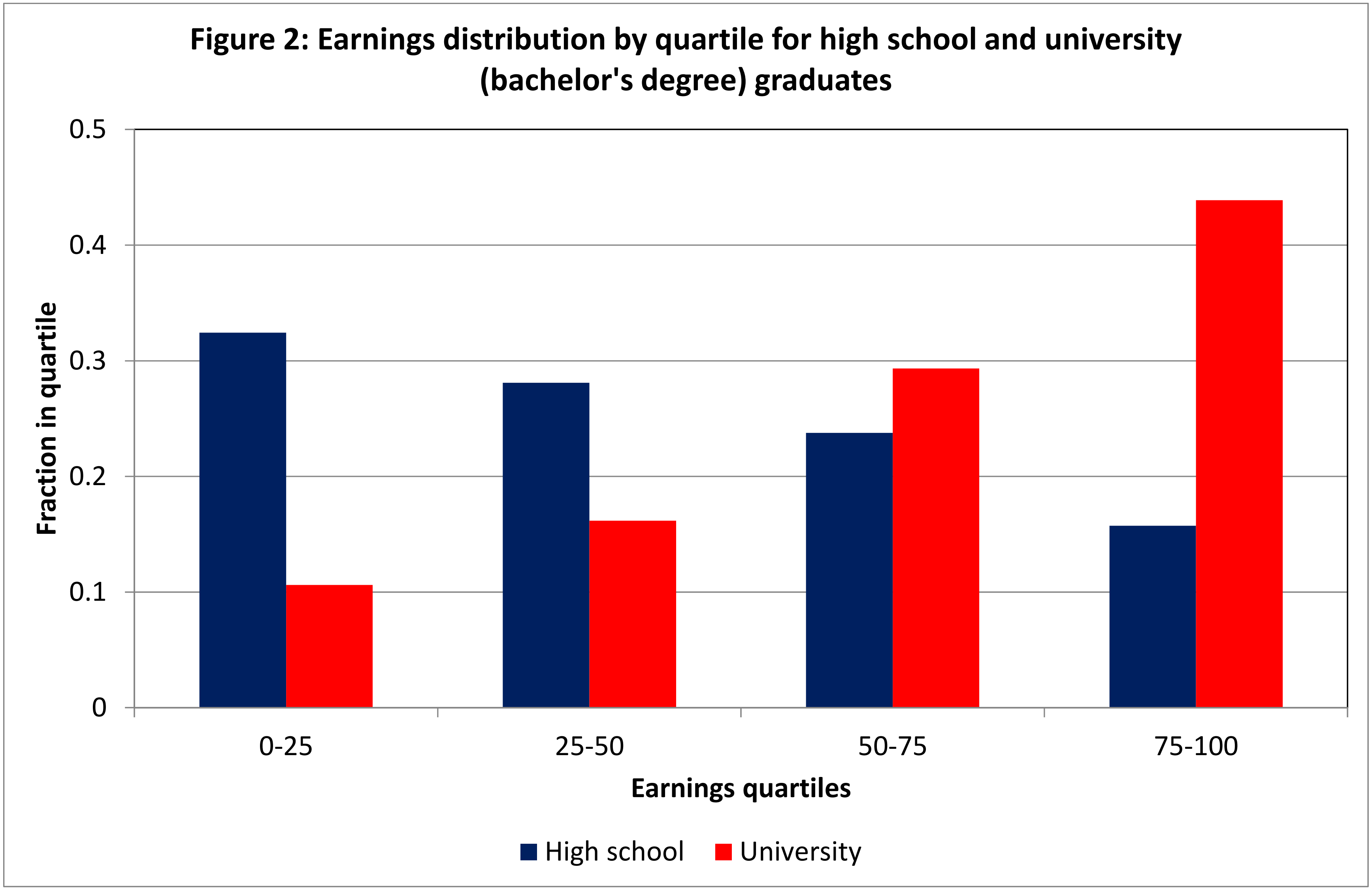
Figure 2: Earnings distribution by quartile for high school and university (bachelor's degree) graduates
0.5
0.4
0.3
0.2
0.1
0
Fraction in quartile
0-25
25-50
50-75
75-100
Earnings quartiles
High school
University

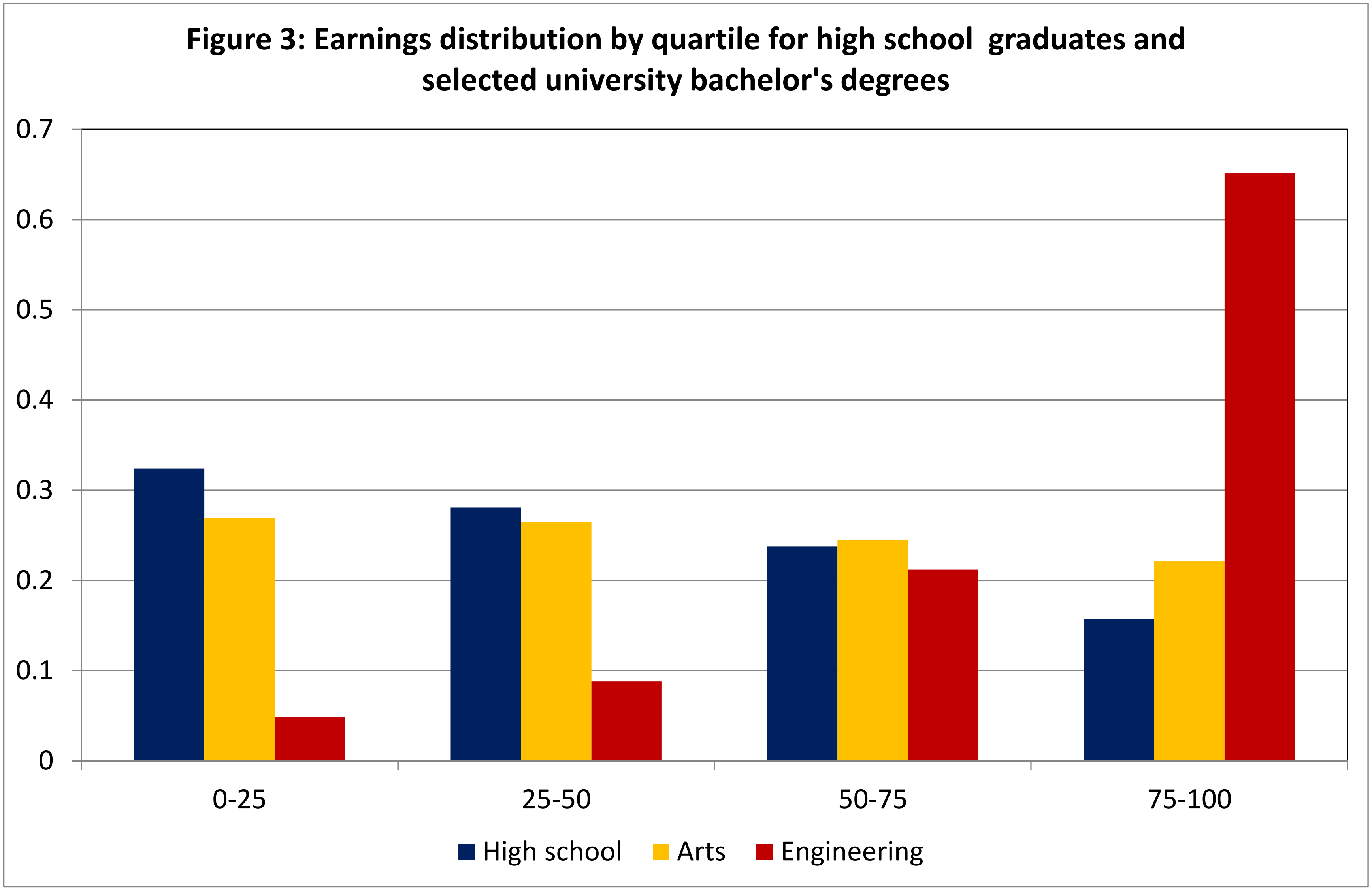
Figure 3: Earnings distribution by quartile for high school graduates and selected university bachelor's degrees
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
0-25
25-50
50-75
75-100
High school
Arts
Engineering

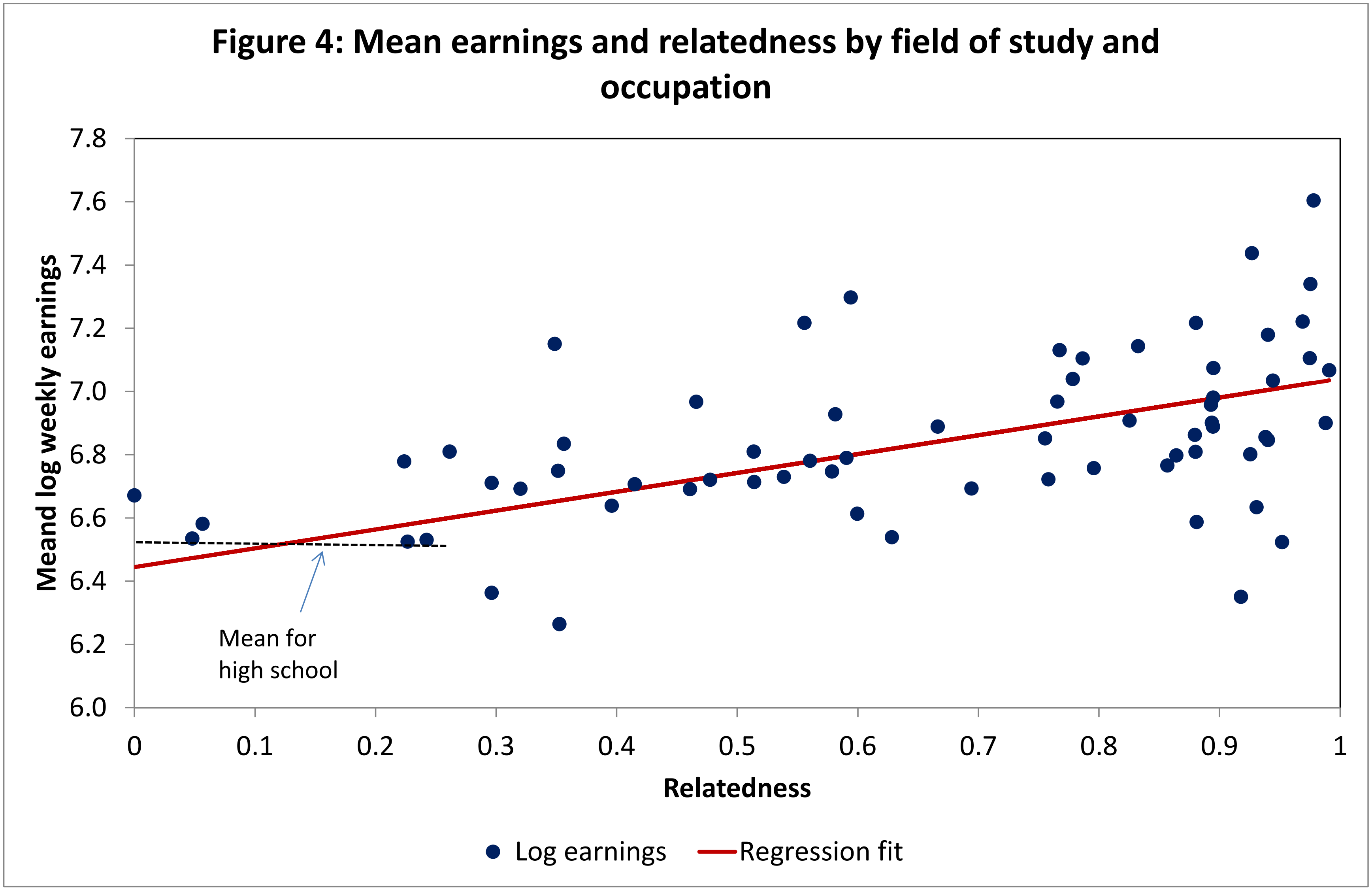
Figure 4: Mean earnings and relatedness by field of study and occupation
7.8
7.6
7.4
7.2
7.0
6.8
6.6
6.4
6.2
6.0
Meand log weekly earnings
Mean for
high school
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
Relatedness
Log earnings
Regression fit

Table 1: Descriptive Statistics, 2005 NGS

	Field of study										All
	Business	Physical & life sc	Computer & math	Engineer. & related	Health, fitness	Education	Soc. sc. & law	Humani-ties	Arts & Comm.	Other	
	A. Percentage of individuals in each field of study										
All	22.5	4.0	4.4	11.0	15.0	9.8	16.6	6.6	4.6	5.5	100
Men	22.2	3.8	8.7	23.3	7.1	6.0	10.8	6.1	4.3	7.7	38.7
Women	22.7	4.1	1.7	3.1	20.0	12.2	20.3	6.9	4.9	4.1	61.3
College	23.8	0.9	4.4	16.8	20.2	2.4	9.8	2.7	7.7	11.5	30.2
University	21.1	5.0	4.3	7.9	13.6	12.5	20.5	8.4	3.8	3.0	54.7
Grad. School	25.2	6.5	5.0	10.5	9.6	14.8	16.3	7.8	1.5	2.7	15.1
	B. Average characteristics by field of study										
% w/ job strongly related to educ	60.9	51.1	64.8	64.7	80.5	83.7	48.7	29.9	40.4	51.3	60.9
% str./somehow related	86.8	73.1	89.9	88.4	92.3	94.6	71.6	51.4	59.6	72.8	81.4
% overqualified	34.9	35.4	30.3	26.3	19.3	14.9	38.0	47.0	42.1	36.3	31.2
Average income	47300	37400	45500	50000	45800	46500	37500	33200	29200	39200	43000

Note: Based on a sample of 10,925 observations from the 2007 follow-up to the 2005 NGS.

Table 2: Distribution of occupation by field of study, 2005 NGS, all diplomas

	Occupation:										
	Manag.	Business	Science	Trades	Manuf.	Health	Soc. Sc. & educ.	Arts	Sales	Proportion in field:	Duncan index:
Field of study:											
Business	**14.8**	**48.8**	4.2	0.8	1.5	0.8	14.0	1.4	13.7	22.5	37.8
Physical & life sc.	3.6	11.9	22.0	1.7	1.3	10.5	37.7	1.7	9.7	4.0	21.3
Math, computer & info.	3.1	11.2	**62.5**	1.5	0.8	0.0	9.6	6.1	5.3	4.5	51.6
Engineering & related	6.3	5.2	**52.1**	**17.1**	**7.3**	0.6	4.7	0.8	6.0	11.0	58.7
Health & fitness	2.2	6.3	1.2	0.3	0.5	**72.7**	9.1	2.0	5.8	15.0	60.4
Education	5.6	1.9	0.6	0.1	0.4	0.9	**85.5**	0.7	4.5	10.1	59.5
Social sc. & law	6.4	21.0	3.2	1.0	0.7	2.1	**42.9**	**8.3**	14.5	16.5	23.9
Humanities	6.4	19.7	4.4	2.8	3.2	1.8	32.1	**9.6**	20.1	6.6	19.6
Arts & communications	5.3	13.1	5.4	3.0	3.6	1.7	8.8	**31.7**	**27.4**	4.3	43.3
Other	5.7	12.8	13.3	**6.4**	**7.9**	1.8	13.8	1.3	**37.1**	5.5	34.3
Proportion in occupation:	7.2	19.7	12.4	3.1	2.3	12.3	25.9	4.5	12.6		

Note: Occupation-field of study cells for which the proportion of workers in the occupation is more than twice as high as in the marginal distribution (for all fields of study combined) are highlighted in bold characters.

Table 3: Job closely or somehow related to education: all post-secondary diploma (NGS)

	Occupation:									
	Manag.	Business	Science	Trades	Manuf.	Health	educ.	Arts	Sales	Total
Field of study:										
Business	**0.976**	**0.895**	0.833	0.461	0.579	0.875	0.940	0.694	0.666	0.868
Physical & life sc.	0.594	0.296	0.895	0.393		0.895	0.881		0.261	0.731
Math, computer & info.	0.927	0.766	**0.975**	0.381			0.944	0.931	0.356	0.899
Engineering & related	0.978	0.582	**0.969**	**0.880**	**0.778**		0.893		0.466	0.884
Health & fitness	0.786	0.591	0.889			**0.991**	0.938	0.918	0.539	0.923
Education	0.881	0.396	0.677			0.880	**0.988**		0.628	0.946
Social sc. & law	0.556	0.560	0.825	0.000	0.112	0.514	**0.940**	**0.857**	0.351	0.716
Humanities	0.349	0.320	0.224	0.057	0.048	0.454	0.864	**0.880**	0.227	0.514
Arts & communications	0.514	0.242	0.758	0.493	0.296	0.172	0.796	**0.952**	**0.352**	0.596
Other	0.767	0.478	0.894	**0.415**	**0.600**	0.795	0.926	0.688	**0.755**	0.728
Total	0.830	0.718	0.920	0.657	0.565	0.959	0.943	0.871	0.506	0.814

Note: Occupation-field of study cells for which the proportion of workers in the occupation is more than twice as high as in the marginal distribution (for all fields of study combined) are highlighted in bold characters.

Table 4: Average log earnings (relative to the mean): all post-secondary diploma (NGS)

	Occupation:									
	Manag.	Business	Science	Trades	Manuf.	Health	Soc. Sc. & educ.	Arts	Sales	Total
Field of study:										
Business	**0.377**	0.078	0.430	0.032	0.161	0.021	0.182	-0.398	-0.232	0.100
Physical & life sc.	0.226	-0.372	0.111	0.082		-0.222	-0.199		-0.543	-0.171
Math, computer & info.	0.430	0.032	**0.200**	-0.246			-0.124	0.001	-0.567	0.092
Engineering & related	**0.465**	0.025	**0.270**	0.091	**0.276**		-0.052		-0.202	0.194
Health & fitness	0.405	-0.278	0.122			**0.161**	-0.059	-0.426	-0.507	0.068
Education	0.749	-0.050	0.134			0.117	**0.148**		-0.536	0.143
Social sc. & law	0.061	-0.109	-0.006	-0.035	-0.004	-0.282	-0.109	-0.180	-0.427	-0.147
Humanities	-0.007	-0.332	0.090	-0.258	-0.411	-0.268	-0.174	-0.082	-0.687	-0.284
Arts & communications	-0.170	-0.459	-0.197	-0.038	-0.295	-0.301	-0.297	**-0.265**	-0.712	-0.387
Other	0.234	-0.080	0.158	0.058	-0.195	-0.025	-0.045	-0.032	-0.221	-0.075
Total	0.319	-0.035	0.219	0.047	0.018	0.123	0.002	-0.210	-0.400	0.000

Note: Occupation-field of study cells for which relatedness exceeds 95 percent are highlighted in bold characters.

Table 5: Regression models for annual income, 2005 NGS

	OLS					IV
	[1]	[2]	[3]	[4]	[5]	[6]
Relatedness	---	---	0.307	0.305	0.302	0.346
			(0.027)	(0.027)	(0.028)	(0.081)
Female	-0.165	-0.161	-0.159	-0.159	-0.157	-0.159
	(0.017)	(0.017)	(0.016)	(0.016)	(0.016)	(0.016)
Bachelor's	0.287	0.254	0.258	0.261	0.260	0.258
	(0.030)	(0.033)	(0.032)	(0.032)	(0.032)	(0.031)
Grad program	0.478	0.422	0.425	0.431	0.428	0.426
	(0.048)	(0.047)	(0.045)	(0.044)	(0.045)	(0.044)
Field of study:						
Business	0.235	0.214	0.138	0.135	0.245	0.128
	(0.064)	(0.022)	(0.031)	(0.031)	(0.005)	(0.043)
Physical & life sc.	-0.106	-0.163	-0.161	-0.160	-0.165	-0.160
	(0.086)	(0.041)	(0.035)	(0.036)	(0.012)	(0.035)
Math, computer,	0.185	0.055	0.003	-0.002	-0.109	-0.004
and information	(0.097)	(0.041)	(0.033)	(0.032)	(0.010)	(0.035)
Engineering	0.337	0.187	0.120	0.115	-0.027	0.111
& related	(0.057)	(0.034)	(0.032)	(0.032)	(0.011)	(0.039)
Health & fitness	0.264	0.115	0.075	0.073	0.060	0.070
	(0.093)	(0.051)	(0.045)	(0.045)	(0.002)	(0.046)
Education	0.209	0.199	0.165	0.164	0.187	0.161
	(0.055)	(0.025)	(0.024)	(0.025)	(0.005)	(0.023)
Humanities	-0.167	-0.155	-0.105	-0.103	-0.092	-0.099
	(0.089)	(0.031)	(0.023)	(0.024)	(0.008)	(0.025)
Arts & communications	-0.116	-0.107	-0.087	-0.085	-0.166	-0.084
	(0.107)	(0.046)	(0.034)	(0.034)	(0.006)	(0.033)
Other	0.176	0.195	0.151	0.151	0.011	0.145
	(0.059)	(0.073)	(0.053)	(0.053)	(0.004)	(0.051)
Controls for:						
Occupations	No	yes	yes	yes	yes	yes
Rank & coop	No	no	no	yes	no	no
Occup * field	---	---	---	---	Control	IV
R-square	0.21	0.26	0.29	0.29	0.30	0.29

Note: Based on 10,925 observations. Standard errors are clustered at the occupation * field of study level. Social sciences and law is the omitted group for field of study.

Table 6: Regressions of Weekly Earnings of Full-time Workers, 2006 Census

	Earnings (Wage and salaries & self empl.) with experience controls						Earnings, age controls	Wage & Sal., exper controls
	All	All	All	Men	Women	Young		
	[1]	[2]	[3]	[4]	[5]	[6]	[7]	[8]
Male	-0.263	-0.231	-0.231			-0.122	-0.252	-0.227
	(0.044)	(0.029)	(0.029)			(0.022)	(0.031)	(0.028)
Field of study:								
Business	0.633	0.260	0.284	0.260	0.325	0.419	0.208	0.331
	(0.083)	(0.043)	(0.057)	(0.053)	(0.067)	(0.046)	(0.056)	(0.045)
Physical & life sc.	0.523	0.332	0.302	0.266	0.342	0.351	0.227	0.325
	(0.087)	(0.048)	(0.054)	(0.055)	(0.054)	(0.046)	(0.056)	(0.046)
Math, computer,	0.651	0.303	0.342	0.307	0.486	0.412	0.264	0.367
and information	(0.061)	(0.045)	(0.047)	(0.042)	(0.047)	(0.031)	(0.049)	(0.039)
Engineering	0.743	0.385	0.422	0.411	0.526	0.517	0.344	0.465
& related	(0.071)	(0.043)	(0.049)	(0.045)	(0.057)	(0.037)	(0.051)	(0.040)
Health & fitness	0.651	0.286	0.295	0.182	0.363	0.423	0.225	0.334
	(0.090)	(0.036)	(0.049)	(0.060)	(0.046)	(0.042)	(0.048)	(0.042)
Education	0.500	0.253	0.270	0.216	0.294	0.461	0.200	0.296
	(0.069)	(0.031)	(0.038)	(0.041)	(0.041)	(0.043)	(0.037)	(0.038)
Social sc. & law	0.573	0.405	0.394	0.346	0.418	0.404	0.316	0.368
	(0.097)	(0.044)	(0.043)	(0.047)	(0.041)	(0.029)	(0.043)	(0.028)
Humanities	0.378	0.334	0.247	0.154	0.317	0.310	0.176	0.261
	(0.089)	(0.046)	(0.039)	(0.056)	(0.043)	(0.026)	(0.039)	(0.030)
Arts & communications	0.189	0.119	0.085	0.014	0.14	0.148	0.008	0.103
	(0.071)	(0.042)	(0.025)	(0.035)	(0.041)	(0.032)	(0.025)	(0.028)
Other	0.415	0.175	0.218	0.174	0.265	0.358	0.142	0.242
	(0.080)	(0.035)	(0.048)	(0.054)	(0.038)	(0.046)	(0.046)	(0.037)

Table 6 (continuation)

	Earnings (Wage and salaries & self empl.) with experience controls						Earnings, age controls	Wage & Sal., exper Controls
	All	All	All	Men	Women	Young		
	[1]	[2]	[3]	[4]	[5]	[6]	[7]	[8]
Relatedness index		0.494						
		(0.098)						
Relatedness dummy			0.182	0.159	0.181	0.131	0.173	0.116
			(0.054)	(0.046)	(0.058)	(0.040)	(0.055)	(0.037)
Occupations	no	yes	yes	yes	yes	yes	yes	Yes
R2	0.24	0.31	0.31	0.28	0.32	0.32	0.26	0.33
Observations	100321	100321	100321	54248	46073	29763	92603	94976

Note: All models also include controls (dummies for 5-year groups) for years of experience, except for the model in column 7 that includes age controls instead. Young workers (column 6) consists of university graduates (bachelor's degree only) age 25-34, and high school graduates age 20-29. In both cases this corresponds to about 3 to 12 years of potential labour market experience. The relatedness index is the fraction of individuals in a given field of study/ occupation cell with a job strongly or somehow related to their education (from the 2005 NGS). The relatedness dummy is a dummy variable equal to 1 in field of study/occupation cells with a relatedness index higher than average, and zero otherwise. The occupation controls (models 2-8) are a set of 24 occupation dummies. Standard errors are clustered at the field of study / occupation level.

Table 7: Decomposition of the Mean Earnings Gap between University (Bach. Degree) and High School Graduates

	% in field	Total gap	Gap due to Occupations	Gap due to Match	Gap due to "pure" return
Field of Study:					
Education	15.2	0.486	0.054	0.162	0.270
Arts & comm.	3.0	0.177	0.009	0.083	0.085
Humanities	9.7	0.370	0.066	0.057	0.247
Soc. sc. & law	21.2	0.564	0.093	0.078	0.394
Business	20.4	0.626	0.185	0.156	0.284
Phys. & life sc.	5.9	0.515	0.118	0.095	0.302
Math, comp. & info.	4.2	0.650	0.153	0.155	0.342
Engin. & related	9.4	0.747	0.164	0.160	0.422
Health & fitness	8.5	0.632	0.177	0.160	0.295
Others	2.5	0.406	0.114	0.074	0.218
All fields	100.0	0.554	0.119	0.124	0.311
(% of total gap)			21.5	22.3	56.2