Return to college education revisited: Is relevance relevant?

Olga Yakusheva *

Marquette University, PO Box 1881, Milwaukee, WI 53201-1881, United States

A R T I C L E   I N F O

A B S T R A C T

This study examines whether the size of the college earnings premium varies depending on the quality of the match between an individual's degree field and his/her occupation. The study uses the Occupational Information Network (O*NET) to obtain a new measure of the quality of occupational match for a sample of 2268 young adults with post-secondary degrees from the restricted use High School and Beyond (1980/92) data. The study finds that people whose occupations better match their degree fields earn significantly higher returns to post-secondary schooling. This result is robust to controlling for an extensive set of pre-existing differences among individuals, and to accounting for differences in earnings across post-secondary degree fields.

1. Introduction

The return to schooling, or the earnings premium associated with an additional year of education, has been a focus of research in economics for decades (see Card, 1999, for a survey of the literature). Among various factors that can affect the return to schooling, occupational mismatch, traditionally defined as over-education, has received a fair amount of attention; however, a consensus as to whether the effect is positive or negative has yet to be reached[1]. While a number of problems have been noted with defining occupational mismatch as *over-education*, or an excess amount of schooling compared to what would normally be required to perform the job, this definition has dominated occupational mismatch literature. The distinct feature of this paper is that it defines occupational mismatch as a situation in which a position is filled by an individual with

a wrong type, rather than an excess amount, of schooling. We take an original (albeit somewhat time consuming) approach involving the use of the on-line Occupational Information Network (O*NET)[2] to establish, one by one, the quality of the match between the field of post-secondary schooling and occupations for 2268 people from the High School and Beyond survey (HS&B). Using this new quality of the match variable, we find that people whose occupations better match their degree fields earn a significantly higher return to schooling (Fig. 1).

There has been a fair amount of skepticism over the assertion that over-education is a valid indicator of occupational mismatch. Ideally, the term *occupational mismatch* should describe the situation in which "workers have more skills and abilities than their jobs utilize" (Clogg & Shockey, 1984; Halaby, 1994). While the traditional definition of occupational mismatch, based on discrepancy between the amount of schooling that is usually required for a job and the amount of schooling the person filling the position has (Sullivan, 1978), clearly tries to follow this conceptualization (Clogg & Shockey, 1984), it has been argued

---

* Tel.: +1 414 288 3409; fax: +1 414 288 5755.
  E-mail address: olga.yakusheva@mu.edu.

[1] Some researchers find that over-qualified workers generally earn a lower return to schooling than equally educated individuals working in jobs that require their degree (Rumberger, 1987; Sicherman, 1991; Verdugo and Verdugo, 1989). Others suggest that over-education may lead to a higher return to schooling by allowing access to higher level positions (Shockey, 1989).

[2] The Occupational Information Network (formerly known as the Dictionary of Occupational Titles) is a comprehensive database of occupational characteristics in the US http://online.onetcenter.org/.
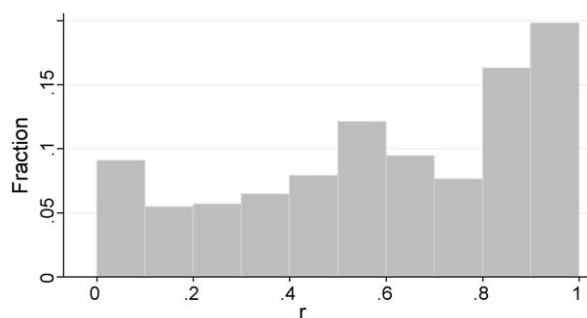
Fig. 1. Frequency distribution of the relevance variable.

that this definition may be too general. In particular, the underlying assumption that workers with more schooling than what is required for their occupations have an excess of *useful* skills may be questionable (Smith, 1986). Moreover, empirical research has found the connection between nominal over-education and the concept of skill mismatch to be "weak and inconsistent" (Halaby, 1994). In the spirit of this criticism, this paper represents the first attempt to examine the relationship among schooling, earnings and occupational mismatch, using a definition of occupational mismatch based on the field of knowledge.

The layout of the paper is as follows. Section 2 discusses the methodological approach of this paper. The HS&B data used in the analysis and the O*NET data on quality of occupational match are described in Section 3. Section 4 presents and discusses the empirical estimates of the relationship among earnings, schooling, and occupational mismatch. Finally, Section 5 provides a summary of findings and our conclusions.

2. Methods

In his influential study, Mincer (1974) developed a relationship between human capital and earnings which, for over three decades, has been used to estimate the return to schooling. Mincer's human capital earnings function is:

$$\text{Model 1} : \log(y_i) = \alpha + \beta S_i + \gamma_1 X_i + \gamma_2 X_i^2 + \Phi' Z_i + u_i. \quad (1)$$

Here, y_i represents labor market earnings, S_i is the number of years of formal schooling, X_i is experience, Z_i is a vector of variables reflecting individual-specific controls, and u_i is an independent identically distributed error term. The coefficient of the schooling variable, β, represents the average conditional earnings premium associated with an additional year of schooling, and is called the return to schooling.

A less restrictive version of model 1, namely one that does not assume linearity in the relationship between schooling and earnings, can be formulated by replacing the continuous schooling variable S with one or more indicator variables that represent various degree levels (typically, an associate degree and/or a bachelor's degree). For example, the relationship between schooling and the log of earnings can be specified as either a continuous linear relationship as in Model 1, or as a non-parametric relationship in the

following way:

$$\text{Model 1} : \log(y_i) = \alpha + \beta BS_i + \gamma_1 X_i + \gamma_2 X_i^2 + \Phi' Z_i + u_i. \quad (1')$$

Here, BS_i is a 0/1 variable equal to one if a person has a bachelor's degree. When the schooling variable is defined as a degree indicator, rather than a continuous variable, the coefficient on the schooling variable, β, reflects a percentage difference in earnings between people with a bachelor's degree and the reference category of people without a 4-year college degree. This coefficient is often referred to as the 4-year college premium.

Mincer's human capital earnings function is derived in part from the assumption that schooling raises the marginal product of labor, which in a competitive market raises the wage. This implies that the schooling variable in the human capital earnings regression should represent the amount of training received by a worker in the tasks performed by the worker in the production process. In reality, however, individuals are often employed in occupations that are less than perfectly matched to their specific training (see Clogg & Shockey, 1984; Manacorda & Petrongolo, 1999; Vahey, 2000). Individuals who work in occupations that do not fully utilize their specific training may experience fewer benefits from their human capital in the workplace, resulting in smaller wage gains than otherwise could be expected.

This paper tests the hypothesis that an individual's return to post-secondary schooling will positively depend on the quality of the match between his/her field of post-secondary training and the requirements of his/her occupation. This hypothesis will be tested by estimating both the linear and the non-parametric versions of Mincer's human capital earnings function (Models (1) and (1')) on a group of people with at least some post-secondary schooling, controlling for the quality of occupational match. This control consists of the variable r, which represents the degree of relevance of a person's field of schooling to his/her occupation, and an interaction term between schooling and its relevance:

$$\text{Model 2} : \log(y_i) = \alpha + \beta_1 r_i + \beta_2 S_i + \beta_3 (S_i \times r_i)$$
$$+ \Phi' Z_i + e_i. \quad (2)$$

$$\text{Model 2} : \log(y_i) = \alpha + \beta_1 r_i + \beta_2 BS_i + \beta_3 (S_i \times r_i)$$
$$+ \Phi' Z_i + e_i. \quad (2')$$

The HS&B sample includes young people all of whom entered high school in the same year (and most of whom also graduated in the same year), thus we do not control for labor market experience, or age, in the earnings models.[3] Since we are unable to examine the relationship between earnings and schooling independently of the relationship between earnings and experience, and because in a given age group the amount of schooling is inversely related to

---

[3] When a control for years of tenure on the current job is added, it proves to be important for wage growth; however, its relationship to earnings appears to be independent of match quality and does not affect any of the estimates presented in this paper.

experience, our estimates of the return to schooling are likely to be biased downward (Mincer, 1974).

In Models (2) and (2′), the change in earnings associated with an increase in relevance, β_1 reveals the magnitude of wage differentials between good and bad occupational matches among people with little post-secondary schooling in the linear returns model (Model (2)), and among people with less than 4 years of post-secondary schooling in the non-parametric model (Model (2′)). Coefficient β_2 represents the return to "irrelevant" post-secondary schooling, i.e., the percentage change in earnings that is associated with having an additional year of post-secondary schooling (Model (2)) or with having a bachelor's degree (Model (2′)), among people who entirely failed to match their degree field to an occupation. Finally, the coefficient of the interaction terms between schooling and its relevance, β_3, reflects the difference in returns to post-secondary schooling between people who fully matched their field of their post-secondary education to an occupation, and people who failed to do so. If the quality of the match between the degree field and occupational requirements does indeed positively affect the return to schooling, we expect the coefficient of the interaction terms β_3 to be positive and statistically significant.

It should be noted, however, that finding of a positive and statistically significant coefficient on the schooling–relevance interaction term would not necessarily imply that the quality of occupational match impacts earnings, due to the likely endogeneity of occupational relevance, and also due to potential systematic differences in employability and earnings across degree fields. Firstly, it is likely that the quality of occupational match is determined endogenously. Unobserved characteristics, such as ability and motivation, can affect the degree to which one's occupation matches his or her degree field if higher-ability individuals are more successful at finding jobs that better utilize their specialized training (Miller, 1984). For example, recent work suggests that the size of ability bias in the estimate of returns to schooling can display a significant amount of variation across individuals (Arias, Hallock, & Sosa-Escudero, 2001).

Controlling for possible endogeneity presents a serious challenge whenever research involves non-experimental data. When a valid source of exogenous variation in the independent variable (i.e. an instrument) is unavailable, and a reliable way to identify observations with "similar" unobservable characteristics (i.e. using twins, or siblings) cannot be found, researchers often resort to controlling for observable factors that they believe to be correlated with unobservable differences among individuals causing self-selection. In this paper, we add a wide variety of ability, aptitude, and motivation variables, measured while the respondents were still in high school, as controls for pre-existing unobserved individual characteristics and self-selection. While this approach is limited in its ability to rid the estimates of ability biases, it can be used as a robustness check and therefore provide some valuable insights, given that fully eliminating endogeneity in this study is not feasible.

Adding controls for underlying individual heterogeneity also allows a better insight into the interpretation of the coefficient of "irrelevant" schooling, β_2. According to the human capital model, a degree in a field that is completely irrelevant to an occupation should not increase a worker's productivity; therefore, the effect of post-secondary schooling on earnings among people who work in occupations for which their degrees are completely irrelevant should be statistically indistinguishable from zero. However, there are two reasons why people can earn positive returns to an irrelevant degree. First, according to Spence (1973), post-secondary education, however irrelevant, may be taken by the employer as a signal of unobserved productivity-enhancing personality traits (also known as the "sheepskin effect" after Hugerford & Solon, 1987, and in Jaeger & Page, 1996). Second, it is possible that through post-secondary education one accumulates general human capital like computer literacy, critical thinking, and social skills, which can augment one's productivity even in occupations unrelated to the specific field or major of the individual's degree. Examining the robustness of the coefficient of the stand-along schooling variable to the inclusion of controls for pre-existing unobserved heterogeneity will help us to discern the extent to which the correlation between "irrelevant" schooling and wages can be attributed to underlying individual differences (the signaling model) rather than to general knowledge acquired through post-secondary schooling (the human capital model).

The second reason why a positive relationship between earnings and relevance should be interpreted with caution, as we mention earlier, is the possibility that degrees in some fields may inherently be less marketable, leading to lower earnings and a lower probability of employment. For example, according to the O*NET, the occupation "Historians" has average annual earnings of $50,790 and projected employment growth of 7–13%, compared to the occupation "Environmental Engineers" that earns on average $72,350 and is expected to grow at a rate exceeding 21%. Thus, a statistical association between earnings and occupational match quality could be a result of earnings and employability differentials across degree fields, and not evidence of the negative effect of occupational mismatch on earnings. In order to examine the role of occupational mismatch independent of systematic differences in degree marketability, we control for earnings differentials across degree fields that exist in the absence of occupational mismatch, using information on earnings by degree field and match quality from both the HS&B and the O*NET. We also estimate our models by controlling for any differences among degree fields, including marketability, with a full set of 2-digit degree field indicators.

In summary, this paper examines the importance of occupational matching for labor market earnings by estimating two versions of Mincer's earnings model with a control for the occupational relevance of an individual's schooling and an interaction term between schooling and its relevance. An extensive set of ability, aptitude, and motivation variables is used to examine the robustness of the findings to controlling for unobserved heterogeneity. Measures of earnings differences across degree fields and degree field fixed effects are used to account for differences in the marketability of different types of post-secondary

Table 1
Comparison between HS&B sample and CPS 1991 sample.

	HS&B	CPS 1991
Observations	2268	2168
Weighted % White	80.0	84.8
Weighted % Female	50.7	48.7
Weighted average annual earnings (x $1000 in 1991)	25.1 (15.55)	24.9 (13.99)
Weighted % with a 4-year college degree	49.6	49.4
% Part-time	n/r	15.8
Weighted average weekly full-time hours	n/r	40.0 (10.8)
Weighted average weekly part-time hours	n/r	27.2[a] (12.1)
Weighted average weekly hours, males	n/r	43.3 (8.9)
Weighted average weekly hours, females	n/r	38.4[b] (7.8)
Weighted average weekly hours, married females with children	n/r	37.7[b] (8.4)

Notes: numbers shown are weighted using corresponding survey weights. Standard errors of the means are in parenthesis. None of the differences between the two samples are statistically significant at the 0.05 level. "n/r" means data not reported.

[a] Different from full-time hours at $p < 0.0001$.
[b] Different from male work hours at $p < 0.001$.

training. With sensible controls for individual heterogeneity and earnings differences among degree fields, the coefficients of the interaction terms between schooling and its relevance can provide an idea of how important occupational match quality may be in determining individual earnings and the return to schooling.

3. Data

The study sample is obtained from the restricted use High School and Beyond (HS&B) data. This survey was conducted by the U.S. Department of Education and contains data on a diversified representative sample of close to 15,000 1980 sophomores and 12,000 1980 seniors. Follow-up surveys were conducted in 1982, when the majority of the respondents finished high school, and later in 1984, 1986, and for the sophomore cohort only, in 1992. The HS&B data are unique in that, during the 1992 follow-up, participants were asked a set of open-ended questions regarding their yearly earnings, industry of employment, occupation, and duties performed on the job, as well as about post-secondary degrees earned and fields or majors in which the degrees were earned, for each year between 1986 and 1992. Due to the fact that the interviews were conducted throughout 1992, the earnings data for 1992 represent an incomplete year of earnings for many respondents. Therefore, this study uses data on earnings, occupation, and schooling from year 1991. Since the senior cohort was not followed-up in 1992, only the sophomore cohort was included in this study.

The study sample includes only high school graduates who went on to receive some post-secondary schooling (e.g. a certificate, a vocational license or technical award, an associate degree, or a bachelor's degree), and who reported yearly wage incomes in 1991 between $1000 and $300,000. Since our study focuses on the link between schooling, occupation, and earnings, we excluded individuals who were self-employed, as well as homemak-

ers, military personnel, and those who were undergoing on-the-job training (research assistants, interns, trainees, etc.). We focused our attention on individuals with at most 4 years of post-secondary schooling because the survey time span (10 years beyond high school graduation) did not allow us to fully observe the transition from school to workplace for individuals with eight or more years of post-secondary education (doctoral and other professional degrees), and because after all the exclusions, there were an insufficient number of individuals with Master's degrees to comprise a stand-alone schooling category. Observations with missing data on gender, race, marital/parenting status, industry of employment and occupation, and post-secondary degree level and field of specialization were deleted. Other missing data (union affiliation status and ability/aptitude/motivation controls) were mean-substituted and missing value indicators were included in all analyses to account for possible non-randomness of missing data. In compliance with the National Center for Education Statistics' restricted data non-disclosure requirements, all sample size numbers in this paper are weighted using normalized survey sampling weights. The final weighted sample size is 2268 individuals.

Table 1 compares demographic and basic socioeconomic characteristics of our study sample to those of a sample of 2168 individuals from the Monthly Basic Current Population Surveys for 1991 (hereafter, CPS 1991). Inclusion and exclusion criteria for creating the CPS 1991 sample were consistent with those used for our HS&B sample.[4] All

---

[4] The Monthly Current Population Survey uses the regular CPS questionnaire eliciting responses to a standard set of labor force participation and demographic questions. Our CPS 1991 sample pools data from all 12 monthly surveys for 1991. Included are all 28-year-old individuals who reported working during the week preceding the interview, had positive weekly earnings, and had 1–4 years of post-secondary schooling. Self-employed individuals, those who never worked, and those working with no pay were excluded.

statistics in Table 1 are weighted using normalized 1980 survey weights for the HS&B sample and using the CPS Adults Final Weight for the CPS 1991 sample. While the HS&B has a slightly smaller weighed proportion of whites (80% in the HS&B vs. 85% in the CPS 1991) and slightly more females (51% vs. 49%) than the CPS 1991 sample, the differences between the samples are not statistically significant. The differences in the weighted average yearly earnings ($25,130 in the HS&B vs. $24,917 in the CPS, in 1991 dollars) and in the weighted proportion of people with a bachelor's degree (50% in the HS&B vs. 49% in the CPS) are also within the margin of error. Table 2 provides names of variables, their definitions and descriptive statistics for the HS&B sample. All results presented in this paper are obtained by estimating weighted regression models.

The earnings variable used in this study is the logarithm of yearly earnings in 1991. The HS&B data do not include information on the hourly wage, the number of hours worked, or part-time status in the 1992 follow-up.[5] In the CPS 1991 sample, 15.8% of individuals were working part-time, and the average part-time worker worked a little over 27 h per week (vs. 40 h per week among full-time workers[6]). Since work hours can be systematically related to the quality of occupational match, not having this information in the HS&B data represents a significant limitation. We use data on individuals' race, gender, marital status, number of children, union membership status, and 2-digit industry and occupational codes, to control for work hours. In the CPS 1991 sample, a similarly defined set of variables explained 15% of total variation in weekly work hours among the individuals.[7] We examine the robustness of our findings to controlling for unmeasured variation in work hours by stratifying the sample based on demographic characteristics likely correlated with work hours, such as gender and number of children.

The schooling variable in this study captures the amount of post-secondary schooling and it is based on the highest degree achieved by 1991. The HS&B distinguishes between three types of non-graduate post-secondary degrees: Certificate/License/Technical Award (536 individuals), Associate's (519 individuals), and Bachelor's (1213 individuals). For the linear specification of the schooling variable in model 2, people in the Certificate/License/Technical Award category were assigned a value of $S = 1$, Associate's degrees were assigned a value of $S = 2$, and Bachelor's degrees were assigned a value of $S = 4$.[8] For the non-parametric specification of the school-

ing variable, we created two schooling categories: a "4-year degree" category that includes the Bachelor's degrees and a "2-year degree or less" category that encompasses all shorter post-secondary degrees. There are two reasons for pooling the two shorter-term post-secondary degree types, Certificate/License/Trade Award and Associate's, into one "2-year degree or less" category. First, given the observation count across the three educational categories, combining the two shorter-term categories builds a more balanced reference group for examining the return to a 4-year college degree. Second, when the two shorter-term degree categories are modeled separately, the resulting regression estimates of earnings differences between the two categories lack statistical significance and power.[9] This could be the result of considerable heterogeneity in the lengths and rigor levels among the post-secondary programs in the Certificate/License/Trade Award category. This category includes all post-secondary programs other than academic programs leading to either an Associate's or a Bachelor's degree, and they range in duration from several weeks to 2 years (and in some cases longer). As such, an unknown and potentially significant number of degrees in this category may be equivalent in terms of duration and amount of human capital they provide to an Associate's degree, making the grounds for dividing shorter programs into two separate categories questionable.

The quality of occupational match was established based on occupational information provided on the U.S. Department of Labor's Occupational Information Network (O*NET).[10] O*NET contains detailed information on a wide variety of occupations, including an array of attributes (skills, abilities, knowledge, etc.) that are required to perform various occupations. The database is constructed by selecting a random sample of businesses that employ workers in a particular occupation, and then surveying a random sample of their employees who work in that particular occupation about their day-to-day experiences on the job. The individual responses are then aggregated to create a variety of numerical and categorical scores that describe various attributes of an occupation. The database is routinely updated.

O*NET classifies different fields of knowledge in 33 broad categories, such as "Administration and Management", "Education and Training", "Biology", "Chemistry", etc. Since HS&B distinguishes between over 100 different post-secondary fields/majors, many O*NET fields of knowledge were matched to more than one HS&B field/major,

---

[5] The hourly wage is available in all of the previous follow-ups, including the 1986 follow-up. However, since the 1986 data were collected only 4 years after most finished high school, the number of individuals with a completed post-secondary degree who were also employed in 1986 is insufficient for reliable statistical analysis.

[6] Full time status as defined as working 35 h or more per week.

[7] The results are available from the author on request.

[8] Note that we could have defined the continuous schooling variable as the total number of years of schooling (i.e. 12 years of secondary schooling plus post-secondary schooling; e.g. 14 years for Associate's degree and 16 years for Bachelor's degree); however this would make it difficult to interpret the coefficients of the stand-alone relevance variable in models

with interaction terms between schooling and relevance because none of the individuals in our sample have less than 12 years of schooling.

[9] Power analysis was based on estimated coefficients and standard errors obtained after estimating a non-parametric specification of Mincer's earning model with an Associate's and a Bachelor's indicators in the equation and the omitted category of Certificate/License/Trade Award. In various specifications of the model, the power of the Associate's degree coefficient estimate varied from 10 to 60%, and the detectable effect sizes for the conventional power level of 80% varied from .25 to .38, clearly exceeding the range of magnitudes of earnings differences between an Associate's degree and the reference category that one could reasonably expect to find. Results are available upon request.

[10] Online Occupational Information Network http://online.onetcenter.org.

Table 2
Variable definitions and descriptive statistics.

Variable	Definition	# Obs	Mean	St. Dev	Min	Max
Variables of interest (1992 survey)						
Earnings	Log of individual annual nominal earnings in 1992	2268	9.98	.59	7.00	12.59
Bachelor degree	=1 if highest degree is a 4-year college degree = 0 if certificate/license/trade award or Associate's	2286	.59	.49	0	1
Schooling	=1 if highest degree is certificate/license/trade award = 2 if highest degree is Associate's = 4 if highest degree is Bachelor's	2268	2.95	1.29	1	4
Relevance	Coefficient of relevance of field of education to occupation in 1992. Continuous variable with range of values from 0 to 1 (0 = irrelevant, 1 = fully relevant)	2268	.61	.31	0	1
Med. relevance	=1 if .2 < relevance ≤ .6	2268	.31	.46	0	1
High relevance	= 1 if .6 < relevance ≤ 1	2268	.55	.49	0	1
Demographic controls (1992 survey)						
White	0 = no, 1 = yes	2268	.80	.47	0	1
Black	0 = no, 1 = yes	2268	.08	.31	0	1
Hispanic[*]	0 = no, 1 = yes	2268	.10	.38	0	1
Gender	0 = male, 1 = female	2268	.51	.49	0	1
Married	0 = never married, widowed, divorced, or separated, in 1992, 1 = married or living together, in 1992	2268	.46	.49	0	1
Children	0 = does not have any children by 1992, 1 = has at least one child by 1992	2268	.29	.45	0	1
Married with children	Interaction term between "Married" and "Children"	2268	.23	.42	0	1
Multiple degrees	=1 if has multiple same-level post-secondary degrees	2268	.08	.27	0	1
Ability controls (1980 survey)						
Locus of control	Continuous measure of self-perceived ability to control life	2268	.77	2.19	−2.41	9
Verbal test score	Standardized item response test score, based on verbal test scores in 1980	2268	0	1	−1.83	2.04
Math test score	Standardized item response test score, based on math test scores in 1980	2268	0	1	−1.94	2.31
Has skills to find out about jobs[**]	Based on question: "do you know how to find out about difference kinds of jobs?" 1 = no, 2 = not sure, 3 = yes	2688	2.60	.60	1	3
Father's education[***]	Highest level of education completed by father, stepfather, or male guardian in 1982	2268	12.91	4.27	0	20

Aptitude controls (1980 survey)						
Has a negative attitude toward planning for the future[**]	Based on question: "How do you feel about the following statement: planning only makes a person unhappy since plans hardly ever work anyway?" 1 = strongly disagree, 2 = disagree, 3 = no opinion, 4 = agree, 5 = strongly agree	2051	3.92	1.02	1	5
Has a positive attitude toward self[**]	Based on question: "How do you feel about the following statement: I take a positive attitude toward myself" 1 = strongly disagree, 2 = disagree, 3 = no opinion, 4 = agree, 5 = strongly agree	2044	4.06	.85	1	5
Thinks of self as overweight	0 = false, 1 = true	2031	.22	.41	0	1
Thinks of self as physically unattractive	0 = false, 1 = true	2007	.14	.35	0	1
Feels depressed[**]	Based on question: "during the past few weeks, did you ever feel depressed or very unhappy?" 1 = never, 2 = once, 3 = several times, 4 = a lot	2059	2.11	.89	1	4
No disciplinary problems at school	0 = false, 1 = true	2031	.87	.33	0	1
Motivation controls (1980 survey)						
Perceived importance of success[**]	Based on question: "How important is being successful at work?" 1 = not important, 2 = somewhat important, 3 = very important	2081	2.88	.34	1	3
Perceived importance of money[**]	Based on question: "How important in your life is having lots of money?" 1 = not important, 2 = somewhat important, 3 = very important	2084	2.23	.59	1	3
Expected level of education[**]	The number of years of post-secondary schooling the respondent is planning to receive	2069	6.35	2.08	1	9
Degree marketability controls (derived)						
HS&B average earnings	For description, see p. 14	2268	10.11	.13	9.02	10.46
O*NET average earnings	For description, see p. 14	2268	10.60	.29	9.58	11.02

Notes: missing value indicators are added to all regressions. The lowest level of father's schooling is "less than high school", and it is coded as 11 years of schooling. Estimation results presented in this paper are robust to coding differences.

[*] Other racial/ethnic categories comprise the remaining .06 of the sample.

[**] Variable is controlled for in regressions with a set of 0/1 indicators, the lowest category is omitted.

[***] The original survey variable is categorical.

Table 3
Cross-reference between HS&B fields/majors and O*NET field of knowledge categories.

O*NET knowledge	HS&B post-secondary education field
Administration and Management	Business Admin/Mngmt, Health/Hospital Admin, Public Admin, Agribusiness/Prod
Biology	Botany, Zoology, Biochem, Biophys, Biostats, Biopsych/Psychobio, Environm. Science
Building and Construction	Construction/Ind. Art
Chemistry	Chemistry, Chemical Engin.
Clerical	Secretarial/Clerical, Other Bus. Supp.
Communications and Media	Communications/PR, Journalism/Broadcast J., Communication Tech
Computers and Electronics	Computer programming, Data Processing, Computer&Info Sci
Customer and Personal Service	Consumer Services, Cosmetology/Hair Styling, Nursing
Design	Architect/Env Design, Graphic/Printing Des.
Economics and Accounting	Accounting, Economics, Finance, Home Econ., Vocational Home Econ.
Education and Training	Early Childhood Educ., Elementary Educ., Secondary Educ., Special Educ., Phys Ed/Health Ed, Education: all other, Child Care/Guidance
Engineering and Technology	Elec/Comm Engin., Chemical Engin., Civil/Ocean Engin., Mechanical Engin., Engin./Sci Tech., Other Engin., Operations Research
English Language	English/Amer Lit., Writing/Create&Tech, Letters: all other
Fine Arts	Fine Art/Art History, Music, Film Arts, Drama/Speech, Writing Create&Tech, Textile/Fashion
Food Production	Nutrition/Food Sci
Foreign Language	Foreign Lang
Geography	Geography, Geology/Earth Sci, Area Sci
History and Archeology	History, Anthropology/Archaeology, Art History, American Civ/Studies
Law and Government	Law, Para-legal/Pre-law, Protective Services, Public Admin., Sociology/Demogr/Criminology, Political Sci, International Rels.
Mathematics	Mathematics, Statistics/Biostats
Mechanical	Mechanical Engin, Mechanics Auto/Air, Mech&Rep: Electronics, Mechanics: all other
Medicine and Dentistry	Dentistry, Medicine, Dental/Med Tech, Community/Mental Health, Practical Nursing, Allied Health: Gen&Other, Speech Path&Audiology, Clinical Health Sci, Veterinary Med, Nursing, Health and Hosp Admin, Public Health, Other Health Sci
Personnel and Human Resources	Basic/Personal Skill, Human Res. Mana
Philosophy and Theology	Philosophy, Religious Studies, Theology
Physics	Physics, Physical Sci: all other
Production and Processing	Agribusiness/Production, Operations Research, Precision Production
Psychology	Psychology, Biopsych/Psychobio
Public Safety and Security	Protective Services
Sales and Marketing	Marketing/Distribution
Sociology and Anthropology	Sociology/Demorg/Crimin, Anthropology/Archaeology, American Civ/Studies, Area Studies, Other Ethnic Studies
Telecommunications	Communication Tech
Therapy and Counseling	Social Work, Community/Mental Health
Transportation	Air Transportation, Transportation: all other

based on their names and descriptions. While no standard cross-walk between HS&B fields/majors categories and O*NETS fields of knowledge categories currently exists in the literature, our approach was guided by the following logic: all HS&B categories whose names (in whole or in part) were also included in the name (or a part thereof) of an O*NET category, or were in whole or in part synonymous with it, were matched to the O*NET category. For example, eight HS&B's fields/majors: "Botany", "Zoology", "Biochemistry", "Biophysics", "Biostatistics", "Biopsychology", and "Environmental Science" were matched to O*NET's "Biology" category. Several HS&B's fields/majors were matched to more than one O*NET's field of knowledge; for example, HS&B's "Biochemistry", in addition to being matched to "Biology", is also matched to "Chemistry". As such, a post-secondary degree in biochemistry will be a match for any occupation requiring either biology or chemistry. Cross-references between HS&B's fields/majors and O*NET fields of knowledge categories are shown in Table 3.

For any given occupation, the O*NET database provides the "Importance" and "Level" ratings of each of the 33 fields of knowledge relative to that specific occupation. The "Importance" rating indicates the degree of importance of a particular field of knowledge to an occupation and it varies continuously from 0 to 100. Fields of knowledge with the range of "Importance" coefficients from 80 to 100 are considered "extremely important", from 60 to 80 are "very important", from 40 to 60 are "important", from 20 to 40 are "somewhat important", and from 0 to 20 are "unimportant". These five categories represent the level anchors on likert scale type questions in the O*NET's questionnaire that is used to solicit the knowledge importance data from survey respondents. The "Level" rating indicates the degree to which a particular field of knowledge is required or needed to perform the occupation, and it also varies continuously from 0 to 100; however, no categorical scale is associated with the level rating. While the importance rating is listed for all occupations and fields of knowledge, the level rating is often missing for fields of knowledge

with low importance (in almost 75% of the cases where a field of knowledge is rated as "unimportant" for an occupation, the level rating is listed as "no data available"). Where both ratings are available, the overall cross-occupational correlation between the two scales is 0.947. Due to the incompleteness of the "level" scale, we use the "importance" scale to establish the relevance of a particular type of education to an occupation.

For everyone in our sample, the individual's self-reported occupation in 1991 was typed into the O*NET's interactive occupation search module, and the resulting report was used to find the importance rating of the individual's degree field to the occupation. For people with multiple post-secondary degrees, if the degrees were consecutive (like Associate and Bachelor), the field of the highest degree was used to determine relevance. For people with multiple non-consecutive same-level degrees (slightly over 8% of the sample), the respondent's answers to the 1992 questions "What is your highest degree?" and "What is the field of the degree?" were used to identify the degree to be matched to the occupation. Since having multiple same-level degrees could be systematically related to unobserved individual characteristics that also relate to schooling, earnings, and occupational relevance, a multiple-degree indicator was included as a control variable in all models.

The retrieved importance coefficients were scaled down to a [0,1] range and used as a proxy variable for occupational relevance of schooling. The sample mean of the relevance variable is .61 (see Table 2). The distribution of the sample values of the relevance variable is left-skewed, with almost one-third of all the matches between degree field and occupation having a rating of .8 or higher. For 330 individuals, their post-secondary degree was listed as being unimportant to the occupations they held, and among them about a quarter had an importance rating of 0.

The importance scale provides an ordinal, rather than cardinal measure of the relevance of a field of knowledge to an occupation. In other words, for any given occupation, fields of knowledge with a relevance value of .2 are more relevant than fields of knowledge with a relevance value of .1, but not necessarily twice as relevant. As an alternative approach to the continuous measure of relevance, and as a robustness check, we convert the continuous relevance variable into a discrete variable with three categories: "Low Relevance", for values of relevance in the range from 0 to .2 which are categorized as "Unimportant" in the O*NET ($n = 330$); "Medium Relevance", for values between .2 and .6 that are considered "Somewhat Important" and "Important" ($n = 736$); and "High Relevance", for values exceeding .6 that the O*NET categorizes as "Very Important" and "Extremely Important" ($n = 1202$). The mean of the continuous relevance variable is .07 in the low relevance category, .42 in the medium relevance category, and .83 in the high-relevance category, and there are no significant differences in the means across schooling levels in any of the categories. These ordinal categories will be used to estimate a non-parametric form of the relationship between earnings, schooling, and relevance.

The set of controls for unobserved individual heterogeneity includes controls for ability (locus of control score,

standardized values of 8th grade math and vocabulary test scores, whether or not the individual has skills to choose a high school program, apply to college, and find a job, and parental level of schooling), controls for aptitude (views on importance of planning for the future, the perception of self-worthiness, being prone to depression, perceiving themselves as being overweight or physically unattractive, and having behavioral problems), and finally controls for motivation (value placed on having a successful career and high income in the future, and expected educational achievement). In order to ensure that individual variation in the ability/aptitude/motivation variables reflects pre-existing individual differences and is not a direct result of differences in post-secondary educational experiences, the controls are taken from the first wave of the survey administered during the sophomore year of high school.

Finally, we try three approaches to control for earnings differences across post-secondary degree fields. In the first approach, we create the O*NET Earnings By Field variable, using the on-line occupational database to gather information on average annual earnings in all occupations for which a given degree field is ranked as "extremely important" (the importance rating is 80 and higher), and compute a log-transformed weighted average of earnings (weighted by the share of each occupation in the sum of employment levels across all occupations for which the field is important). The second approach creates the HS&B Earnings By Field variable from our study sample as the log-transformed conditional mean of earnings of individuals who fully matched their post-secondary degrees to their occupations ($r \geq .8$), for each given post-secondary degree field. Since both variables reflect earnings across degree fields at a high level of quality of occupational matching, their values, on average, exceed the actual earnings of individuals (in Table 2, the mean of the O*NET Earnings By Field variables is 10.60 and the mean of the HS&B variable is 10.11, compared to the mean of earnings of 9.98 in our sample). The O*NET Earnings By Field variable has more variability than the HS&B variable (the standard deviations are .29 and .13, respectively), and the correlation coefficient between the two variables is low, .05. Lastly, in the third and most general approach, we create a set of 2-digit indicators based on HS&B degree fields.

4. Results

4.1. Schooling, relevance, and earnings

Tables 4A and 4B show estimates of the return to post-secondary schooling with and without controls for the quality of the match between occupation and degree field. In Table 4A, schooling is represented as an indicator variable for a 4-year college degree. The reference level of schooling is shorter-term degrees, including vocational licenses/degrees/certificates and associate's degrees. In Table 4B, schooling is modeled as a continuous variable representing the number of years of post-secondary education (up to 4). The first column in both tables is the conventional human capital earnings model that does not account for the quality of the match between a per-

Table 4A
Relationship between earnings, schooling, and relevance using a categorical schooling variable.

	(1)	(2)	(3)	(4)	(5)
4-year degree	0.156 (5.856)***	0.164 (6.276)***	0.037 (0.744)	0.161 (6.137)***	0.047 (0.797)
Relevance		0.301 (8.210)***	0.199 (4.001)***		
(4-year degree) × (relevance)			0.216 (3.013)***		
Med. relevance				0.062 (1.727)*	0.033 (0.659)
High relevance				0.210 (6.283)***	0.126 (2.763)***
(4-year degree) × (Med. Relevance)					0.067 (0.945)
(4-year degree) × (High Relevance)					0.175 (2.661)***
R-squared	0.236	0.259	0.262	0.256	0.259

Notes: sample size is 2268 individuals with post-secondary degrees. The reference level of schooling is a 2-year degree or less. "Relevance" is a continuous variable taking values from [0,1]. "Med. Relevance" and "High relevance" are 0–1 indicators equal to 1 if degree is "somewhat important" or "important" for "Med. relevance" and "very important" or "extremely important" for "High relevance". The reference level of relevance is "Low relevance". Other controls included in all regressions are gender, race, marital and parenting status, union status, an indicator for having multiple post-secondary degrees, and 2-digit occupation and industry controls. Absolute value of *t*-statistics in parentheses.

* Is significant at 10%.
*** Is significant at 1%.

Table 4B
Relationship between earnings, schooling, and relevance using a continuous schooling variable.

	(1)	(2)	(3)	(4)	(5)
Years of PSE	0.059 (5.726)***	0.062 (6.178)***	0.008 (0.447)	0.061 (6.017)***	0.002 (0.107)
Relevance		0.300 (8.194)***	0.058 (0.727)		
(Years of PSE) × (Relevance)			0.091 (3.388)***		
Med. relevance				0.062 (1.721)*	−0.066 (0.803)
High relevance				0.210 (6.282)***	−0.011 (0.147)
(Years of PSE) × (Med. Relevance)					0.049 (1.800)*
(Years of PSE) × (High Relevance)					0.083 (3.319)***
R-squared	0.236	0.258	0.262	0.256	0.260

Notes: sample size is 2268 individuals with post-secondary degrees. "Years of PSE" is the number of years of post-secondary schooling (1–4). "Relevance" is a continuous variable taking values from [0,1]. "Med. Relevance" and "High relevance" are 0–1 indicators equal to 1 if degree is "somewhat important" or "important" for "Med. relevance" and "very important" or "extremely important" for "High relevance". The reference level of relevance is "Low relevance". Other controls included in all regressions are gender, race, marital and parenting status, union status, an indicator for having multiple post-secondary degrees, and 2-digit occupation and industry controls. Absolute value of *t*-statistics in parentheses.

* Is significant at 10%.
*** Is significant at 1%.

son's occupation and his/her degree field. The estimation results after fitting an earnings model with a control for relevance and an interaction term between schooling and relevance are presented in columns 2 and 3 for the continuous relevance variable, and repeated in columns 4 and 5 for the categorical relevance variable, in both tables.[11]

At this point, the analysis does not control for individual heterogeneity in ability, aptitude, and motivation, nor does it account for differences in earnings across degree fields.

The coefficient of schooling in column 1 of Table 4A suggests that the conditional wage differential between 4-year college graduates and people with shorter-term degrees is 15.6%. Since the average number of years of

---

[11] In all regressions in Tables 4A and 4B, other controls include the individual's race, gender, marital status, whether or not they have a child, whether or not the respondent is a member of a professional union, a set of 2-digit occupational categories, and a set of 2-digit industry controls (see Table 2 for the descriptions of the variables). The HS&B data are incredibly rich, which in principle allows controlling for a much bigger set of covariates. Adding more covariates such as geographical codes, spousal characteristics, the number and ages of children, and others, does not alter

the results qualitatively. The analysis presented in this paper is based on a set of covariates that have been most frequently used in the literature. Due to the design of the survey, there is little variation in the ages of the respondents. Therefore, age is not included as a right-hand side variable in any of the earnings regressions in this study.

post-secondary schooling in the reference category is 1.5 years, this estimate is consistent with the 5.9% earnings premium for an additional year of schooling in column 1 of Table 4B. Both the non-parametric and the linear estimates are statistically significant and their magnitudes are consistent with the lower end of the range of OLS estimates of Mincer's earnings model presented elsewhere in the literature (see Card, 1995, 1999; Jaeger & Page, 1996). The relatively low magnitude of our estimates is likely the result of the confounding effect of experience.

The importance of occupational match as a predictor of individual's labor market earnings is examined in columns 2 through 5 of Tables 4A and 4B. Interestingly, controlling for occupational relevance of schooling slightly increases the return to schooling (.164 vs. .156 in columns 1 and 2 of Table 4A, and .062 vs. .059 columns 1 and 2 of Table 4B). This is due to the fact that in our study sample, individuals with shorter post-secondary degrees are more likely to work in occupations for which their degrees are more directly related than are college graduates. The mean of the relevance variable among people with 2 years or less of post-secondary schooling is .62, compared to .58 among people with 4-year college degrees, and the difference is significant at the .01 level (these numbers are not in the table). It is possible that short-term degrees develop skills that are highly relevant for a specific occupation, whereas 4-year college degrees prepare individuals more generally for a broader range of occupations.

The estimates in Tables 4A and 4B suggest that, on average, people who matched their degree to an occupation ($r = 1$) earn 30% more than people who work in occupations that do not utilize their post-secondary training ($r = 0$) (column 2 in Tables 4A and 4B). Compared to people whose degree fields have low relevance and are considered "unimportant" for their occupations ($r < .20$), people whose degrees have medium relevance ("somewhat important" or "important", $.20 \leq r < .60$) earn roughly 6% more, and those with highly relevant degrees ("very important" and "extremely important", $r \geq .60$) earn 21% more (column 4 in Tables 4A and 4B). We further observe that the relevance earning premium is larger for higher degree levels, as the coefficient of the interaction term between schooling and relevance is positive and significant in all models (columns 3 and 5 of Tables 4A and 4B). For example, estimates in column 3 of Table 4A show that among people with 2 years of post-secondary schooling or less, individuals who matched their degree to an occupation earn 19.9% more than individuals who did not, while the relevance earning premium for people with a 4-year degree is more than twice as high, or 41.5% (.199 + .216 = .415). Similarly, in column 5 of the same table that shows estimates based on categorical specification of the relevance variable, the increase in the earnings premium for having a highly relevant degree is lower among people with at most 2 years of post-secondary schooling, 12.6%, than among people with 4-year degrees 30.1% (.126 + .175 = .301). The estimates of the continuous schooling model in Table 4B support the findings of a positive earnings differential between good and bad occupational matches that increases with more schooling.[12]

A comparison of estimates of the continuous and the categorical relevance variable reveals non-linearities in the relationship between the log of earnings and the importance rating of the degree field for the occupation. In particular, while the overall relevance premium is 30% (the coefficient of the continuous relevance variable in column 2 is .301 and .300 in column 2 of Tables 4A and 4B, respectively), the incremental increase in earnings between low and medium relevance ranges is only marginally statistically significant and is less than one-third the size of the incremental increase in earnings between low and high relevance ranges (.062 vs. .210). Furthermore, the estimates in column 5 of Table 4A show that the disproportional earnings premiums for high levels of relevance exist among both 4-year academic programs and shorter post-secondary programs. For short-term programs, the low to medium relevance earnings premium is 3.3% and the medium relevance to high relevance earnings premium is 12.6%; for a 4-year degree program, the low to medium and the low to high relevance earnings premiums are 10% (.033 + .067) and 30.1% (.126 + .175), respectively. Both of the low to high relevance earnings premiums are statistically significant at the .01 level, and the low to medium premium for a 4-year degree is significant at the .05 level[13]; however, the low to medium relevance earnings premium for shorter degrees is statistically insignificant. Similarly, according to the estimates in the linear return to schooling model in column 5 of Table 4B, there may be no positive earnings differential between medium and low relevance levels among people with low levels of post-secondary schooling,[14] while the low to high earnings premium is 11.4% and it is significant at the .05 level[15]. These findings are consistent with the fact that the O*NET importance scale provides an ordinal, and not a cardinal, measure of the relevance of a degree field to an occupation.

Estimates in columns 3 and 5 of Table 4A show that having an irrelevant post-secondary degree does not significantly increase earnings. While the coefficient of the stand-alone schooling variable is positive in all models, none of the estimates of the return to irrelevant schooling are statistically significant. However, since our estimates of the return to schooling are likely to be biased downward as a result of our inability to control for experience, the fact that the estimates are generally positive suggests that some individuals may benefit financially from their irrelevant degrees.

---

[12] $\partial \log(\text{Earnings})/\partial(\text{Relevance}) = .058 + .091 \times S$, which is equal to .195 for short-term degrees ($S = 1.5$, mean years of schooling among short-term degrees), and .422 for a 4-year college degree ($S = 4$).

[13] F-test of the sum of the two medium relevance coefficients in column 5 of Table 4A, .033 and .067; F-statistic = 3.99, p-value = .046.

[14] $\log(\text{Earnings})|_{S=1.5,\ r=\text{low}} - \log(\text{Earnings})|_{S=1.5,\ r=\text{med}} = -.066 + .049 \times 1.5 = 0.008$; F-statistic = .05, p-value = .735.

[15] $\log(\text{Earnings})|_{S=1.5,\ r=\text{low}} - \log(\text{Earnings})|_{S=1.5,\ r=\text{high}} = -.011 + .083 \times 1.5 = 0.114$; F-statistic = 2.02, p-value = .036.

Table 5A

Impact of adding ability/aptitude/motivation controls on the estimated relationship between earnings, occupational relevance of schooling, and relevance using a categorical schooling variable.

	(1)	(2)	(3)	(4)	(5)
4-year degree	0.091	0.099	−0.022	0.096	−0.025
	(3.132)***	(3.480)***	(0.443)	(3.358)***	(0.411)
Relevance		0.299	0.201		
		(8.211)***	(4.067)***		
(4-year degree) × (Relevance)			0.207		
			(2.933)***		
Med. relevance				0.063	0.026
				(1.790)*	(0.530)
High relevance				0.212	0.126
				(6.405)***	(2.786)***
(4-year degree) × (Med. Relevance)					0.084
					(1.192)
(4-year degree) × (High relevance)					0.180
					(2.769)***
R-squared	0.274	0.296	0.299	0.294	0.297

Notes: sample size is 2268 individuals with post-secondary degrees. The reference level of schooling is a 2-year degree or less. "Relevance" is a continuous variable taking values from [0,1]. "Med. Relevance" and "High relevance" are 0–1 indicators equal to 1 if degree is "somewhat important" or "important" for "Med. relevance" and "very important" or "extremely important" for "High relevance". The reference level of relevance is "Low relevance". Ability/aptitude/motivation controls are included in all regressions. Other controls included in all regressions are gender, race, marital and parenting status, union status, an indicator for having multiple post-secondary degrees, and 2-digit occupation and industry controls. Absolute value of *t*-statistics in parentheses.

* Is significant at 10%.

*** Is significant at 1%.

4.2. Interpretation

In this section, we will explore three reasons, besides human capital underutilization, why the match quality between degree field and occupation may be positively associated with labor market earnings. First, we will examine whether positive selection on unobserved characteristics makes an individual with a high earnings potential more likely to match his/her degree to an occupation, thus creating a spurious positive correlation between relevance and earnings. Second, we will control for earnings differences across degree fields to explore whether some degree fields may be more marketable, leading to higher earnings and an increased likelihood of finding a job. Lastly, we will discuss whether the positive association between earnings and occupational relevance observed in our sample could have been produced by systematic unmeasured variation in work hours, rather than reflecting differences in wages.

Selection on unobserved individual characteristics may play an important role in our analysis. In an attempt to reduce underlying individual heterogeneity – to the extent that conventional regression methods allow us to do so

Table 5B

Impact of adding ability/aptitude/motivation controls on the estimated relationship between earnings, schooling, and occupational relevance of schooling, using a categorical schooling variable.

	(1)	(2)	(3)	(4)	(5)
Years of PSE	0.033	0.036	−0.014	0.035	−0.025
	(2.914)***	(3.242)***	(0.747)	(3.153)***	(1.073)
Relevance		0.298	0.071		
		(8.193)***	(0.891)		
(Years of PSE) × (Relevance)			0.086		
			(3.218)***		
Med. relevance				0.063	−0.075
				(1.786)*	(0.925)
High relevance				0.212	−0.008
				(6.402)***	(0.105)
(Years of PSE) × (Med. Relevance)					0.053
					(1.957)*
(Years of PSE) × (High. Relevance)					0.082
					(3.328)***
	0.274	0.295	0.299	0.293	0.297

Notes: sample size is 2268 individuals with post-secondary degrees. "Years of PSE" is the number of years of post-secondary schooling (1–4) "Relevance" is a continuous variable taking values from [0,1]. "Med. Relevance" and "High relevance" are 0–1 indicators equal to 1 if degree is "somewhat important" or "important" for "Med. relevance" and "very important" or "extremely important" for "High relevance". The reference level of relevance is "Low relevance". Ability/aptitude/motivation controls are included in all regressions. Other controls included in all regressions are gender, race, marital and parenting status, union status, an indicator for having multiple post-secondary degrees, and 2-digit occupation and industry controls. Absolute value of *t*-statistics in parentheses.

* Is significant at 10%.

*** Is significant at 1%.

– we control for an extensive set of variables that capture the respondent's ability, aptitude, and motivation (Tables 5A and 5B). A comparison of estimates in column 1 between Tables 4A and 4B and Tables 5A and 5B shows that selection on unobservable characteristics is responsible for almost half of the positive association between schooling and earnings. While the return to post-secondary schooling remains statistically significant at the .01 level, adding controls for unobserved individual heterogeneity reduced its magnitude by almost half of its original value. For example, the return to a 4-year degree decreased by 6.5 percentage points, from 15.6% to 9.1% (column 1 of Table 5A compared to column 1 of Table 4A), and the average return to an additional year of post-secondary schooling went down by 2.6 percentage points, from 5.9% to 3.3% (column 1 of Table 5B compared to column 1 of Table 4B). Both of the reductions in the magnitudes of the coefficients are statistically significant at the .05 level (not shown in the table).[16]

A further comparison of the estimates of return to schooling between Tables 4A and 4B, and Tables 5A and 5B, reveals that the decrease in the overall return to schooling arises primarily from a reduction in the return to schooling among people who work in occupations for which their schooling is matched poorly. Specifically, comparing estimates of the schooling coefficient in columns 3 and 5 between Tables 4A and 5A shows that the 4-year irrelevant college degree premium decreases by 6–7 percentage points (from 3.7% to negative 2.2% in column 3, and from 4.7% to negative 2.5% in column 5; the differences are significant at the 0.01 level), and the return per year of irrelevant schooling decreases by 2–3 percentage points (from .8% to negative 1.4% in column 3, and from .2% to negative 2.5% in column 5, the differences are significant at the 0.01 level). While none of the negative returns to irrelevant schooling in Tables 5A and 5B are statistically significant, they suggest that once selection on unobserved heterogeneity is controlled for, some individuals with an irrelevant post-secondary degree may be earnings less than individuals without any post-secondary schooling at all, possibly because they have less work experience. The finding that the return to irrelevant post-secondary schooling observed in Tables 4A and 4B shrinks dramatically after controlling for ability, aptitude, and motivation is consistent with the signaling model, suggesting that a college degree is associated with higher earnings in part because it reflects superior underlying labor market characteristics of its holder.

While adding controls for underlying individual heterogeneity reduces the overall return to schooling, the magnitude of the difference in earnings between good and bad schooling-occupational matches remains virtually unaffected. In column 2 of Table 5A, the average difference in earnings between bad and good occupational matches is 29.9% (compared to 30.1% in Table 4A; the p-value of

the difference between the two coefficients is .897).[17] In column 3 of the same table, the relevance earnings premium among people with up to 2 years of college is 20.1%, and it is twice as high among people with 4-year college degrees, 40.8% (vs. 19.9% and 41.5% in Table 4A; the p-values of the coefficient differences between the tables are .768 and .876 for the two schooling levels, respectively). Estimates in column 4 of Table 5A show that the relevance premium continues to be the largest (21.2%) between highly relevant and irrelevant degrees, while the low to mid-relevance earnings premium is only 6.3% (the corresponding estimates in Table 4A were 21.0% and 6.3%, and the p-values of the differences are .875 and .789, respectively). It appears that controlling for unobserved heterogeneity slightly reduces the return to relevance for short-term degrees, while slightly increasing the return to relevance for 4-year college degrees; however, the differences are not statistically significant (p-values of the differences between the coefficients are all higher than .750). Thus, the findings provide little evidence to support the idea that the relationship between earnings and occupational match quality is driven by self-selection.

Next, we will explore the possibility that the positive association between occupational relevance and earnings is a result of the fact that some degree fields may be more marketable than others, leading to a positive correlation between an individual's earnings and likelihood of finding a job for which his/her degree field is highly relevant. To condition out variation in earnings across degree fields in our model, we include an Earnings By Degree Field control that reflects earnings differences across degree fields. In Table 6A, the control is based on earnings data for occupations that are relevant to a given degree field, which we collected from the O*NET occupational information network. In Table 6B, we use HS&B data to control for the earnings of individuals with degrees in a given field who work in occupations for which the degree field is highly relevant. Finally, in Table 6C we add a full set of 2-digit degree field indicators to control for any unmeasured systematic differences across degree fields.

Both the O*NET and the HS&B Earnings By Degree Field variables are positively related to individual earnings. In Table 6A, the coefficient of the Earnings By Field variable, or the conditional elasticity of individual earnings to average earnings by degree field, is around .04 in all models; however, it is not statistically significant (p-values range from .302 to .357). In Table 6B, the estimates of the conditional elasticity are around .33 and they are statistically significant at the .01 level in all models. (These estimates should be interpreted with caution because by construction, the HS&B earnings by field variable is correlated with individual earnings.) In the models that control for differences among degree fields with 2-digit degree field indicators (Table 6C), the degree field indicators are jointly significant at the 0.001 level and contribute about 2% to the R-squared in all models.

Adding controls for earnings differences across degree fields changes the estimates of the relevance earnings pre-

---

[17] See footnote 16.

Table 6A

Impact of controlling for earnings differentials between fields of post-secondary schooling using O*NET Earnings By Field on the estimated relationship between earnings, schooling, and occupational relevance of schooling.

	Schooling = 4-year college degree		Schooling = years of PSE	
	(1)	(2)	(3)	(4)
Schooling	−0.019	−0.021	−0.013	−0.023
	(0.381)	(0.342)	(0.695)	(1.011)
Relevance	0.204		0.076	
	(4.127)***		(0.945)	
(Schooling) × (Relevance)	0.203		0.084	
	(2.872)***		(3.159)***	
Med. relevance		0.032		−0.068
		(0.639)		(0.835)
High relevance		0.131		−0.000
		(2.887)***		(0.004)
(Schooling) × (Med. Relevance)		0.081		0.051
		(1.144)		(1.909)*
(Schooling) × (High Relevance)		0.176		0.081
		(2.693)***		(3.253)***
O*NET Earnings By Field	0.041	0.042	0.038	0.038
	(0.994)	(1.000)	(0.927)	(0.917)
R-squared	0.299	0.297	0.299	0.297

Notes: sample size is 2268 individuals with post-secondary degrees. In the first two columns, "schooling" is an indicator for having a 4-year college degree; in the second two columns, "schooling" is the number of years of post-secondary education. "O*NET Earnings By Field" is the mean of earnings by field of schooling from O*NET. "Relevance" is a continuous variable taking values from [0,1]. "Med. Relevance" and "High relevance" are 0–1 indicators equal to 1 if degree is "somewhat important" or "important" for "Med. relevance" and "very important" or "extremely important" for "High relevance". The reference level of relevance is "Low relevance". Ability/aptitude/motivation controls are included in all regressions. Other controls included in all regressions are gender, race, marital and parenting status, union status, an indicator for having multiple post-secondary degrees, and 2-digit occupation and industry controls. Absolute value of *t*-statistics in parentheses.

* Is significant at 10%.
*** is significant at 1%.

Table 6B

Impact of controlling for earnings differentials between fields of post-secondary schooling using HS&B Earnings By Field on the estimated relationship between earnings, schooling, and occupational relevance of schooling.

	Schooling = 4-year college degree		Schooling = years of PSE	
	(1)	(2)	(3)	(4)
Schooling	−0.011	−0.020	−0.012	−0.024
	(0.224)	(0.327)	(0.643)	(1.072)
Relevance	0.213		0.100	
	(4.338)***		(1.276)	
(Schooling) × (Relevance)	0.169		0.073	
	(2.395)**		(2.721)***	
Med. relevance		0.014		−0.082
		(0.286)		(1.007)
High relevance		0.124		0.002
		(2.764)***		(0.028)
(Schooling) × (Med. Relevance)		0.075		0.049
		(1.070)		(1.829)*
(Schooling) × (High Relevance)		0.155		0.073
		(2.384)**		(2.979)***
HS&B Earnings By Field	0.334	0.331	0.330	0.328
	(5.041)***	(4.971)***	(4.974)***	(4.923)***
R-squared	0.307	0.304	0.307	0.305

Notes: sample size is 2268 individuals with post-secondary degrees. In the first two columns, "schooling" is an indicator for having a 4-year college degree; in the second two columns, "schooling" is the number of years of post-secondary education. "HS&B Earnings By Field" is the mean of earnings of good matches ($r > .8$) by degree level and field of knowledge. "Relevance" is a continuous variable taking values from [0,1]. "Med. Relevance" and "High relevance" are 0–1 indicators equal to 1 if degree is "somewhat important" or "important" for "Med. relevance" and "very important" or "extremely important" for "High relevance". The reference level of relevance is "Low relevance". Ability/aptitude/motivation controls are included in all regressions. Other controls included in all regressions are gender, race, marital and parenting status, union status, an indicator for having multiple post-secondary degrees, and 2-digit occupation and industry controls. Absolute value of *t*-statistics in parentheses.

* Is significant at 10%.
** Is significant at 5%.
*** Is significant at 1%.

Table 6C
Impact of controlling for degree field indicators on the estimated relationship between earnings, schooling, and occupational relevance of schooling.

	Schooling = 4-year college degree		Schooling = years of PSE	
	(1)	(2)	(3)	(4)
Schooling	0.020	0.024	0.002	−0.007
	(0.384)	(0.384)	(0.122)	(0.282)
Relevance	0.159		0.040	
	(3.132)***		(0.494)	
(Schooling) × (Relevance)	0.191		0.079	
	(2.653)***		(2.894)***	
Med. relevance		0.010		−0.090
		(0.189)		(1.09)
High relevance		0.091		−0.035
		(1.979)**		(0.468)
(Schooling) × (Med. Relevance)		0.075		0.051
		(1.064)		(1.863)*
(Schooling) × (High Relevance)		0.163		0.077
		(2.475)**		(3.060)***
F-statistic of degree field indicators	2.68***	2.77***	2.65***	2.74***
R-squared	0.325	0.324	0.325	0.324

Notes: sample size is 2268 individuals with post-secondary degrees. In the first two columns, "schooling" is an indicator for having a 4-year college degree; in the second two columns, "schooling" is the number of years of post-secondary education. "Relevance" is a continuous variable. "Med. Relevance" and "High relevance" are 0–1 indicators equal to 1 if degree is "somewhat important" or "important" for "Med. relevance" and "very important" or "extremely important" for "High relevance". The reference level of relevance is "Low relevance". Ability/aptitude/motivation controls are included in all regressions. Other controls included in all regressions are gender, race, marital and parenting status, union status, an indicator for having multiple post-secondary degrees, and 2-digit occupation and industry controls. Absolute value of *t*-statistics in parentheses.

  * Is significant at 10%.
 ** Is significant at 5%.
*** Is significant at 1%.

miums only slightly in all four specifications of the earnings regression model, and all of the changes are well within the margin of error. The estimates in Table 6C show that even among people with degrees in the same field, there are still large and statistically significant earnings premiums matching the degree to an occupation, for all degree levels. Thus, the idea that individuals with low occupational match quality earn less because they have chosen a degree in a field that pays less and is more difficult to match to an occupation is not supported by the data.

Lastly, since the dependent variable in our analysis is yearly earnings rather than the hourly wage, the estimates of the relationship between occupational relevance and earnings have to be interpreted with caution. While little is known about the link between occupational relevance and work hours, it seems plausible that people who work in occupations that are highly relevant to their degree field may be working more hours than people with jobs that do not utilize their post-secondary training. If there is a positive correlation between work hours and occupational relevance, one would expect to observe positive earnings differentials associated with better occupational match quality, even if occupational mismatch has little impact on productivity or wages.

Since the HS&B data do not include information on work hours, we cannot examine this issue with the degree of thoroughness and rigor it deserves. We used information on work hours in the comparison sample from the Current Population Survey (Table 1) to gain some understanding about the extent to which work hours would have to be correlated with occupational relevance in order to invali-

date our results.[18] We found that in order for the relevance earnings premium of 20.1% among people with up to 2 years of post-secondary schooling to be produced by variation in work hours, the share of part-time workers among people in the low relevance category would have to be at least .57, while the share of part-time workers in the high-relevance category would have to be lower than .13. To negate the relevance earnings premium of 40.8% among college graduates, over 93% of people in the low relevance category, and less than 4% in the high-relevance category would have to be part-time workers. Given that our analysis conditions out a part of systematic variation in work hours by controlling for gender, race, marital and parental status, union status, industry and occupation, and an extensive set of ability, aptitude, and motivation variables, it does not seem likely that the residual variation in work hours has a systematic component large enough to be driving

---

[18] Let l be the proportion of part time workers who have with low levels of occupational relevance, then the proportion of part-time workers who have high levels of occupational relevance is $(.158 − l)$. Let Pl and Ph be the shares of the low-relevance and the high relevance groups in the sample. Given that the ratio of full-time to part-time weekly hours is $40/27.2 = 1.47$, the part-time adjusted earnings ratio of the low relevance group to the high relevance group is: $[1.47l/Pl + (1 − l/Pl)]/[1.47 (.158 − l)/Ph + (1 − (.158 − l)/Ph)]$. To negate the relevance earnings premium, the part-time adjusted earnings ratio would have to be larger than $(1 + \beta1)$ for 2-year degrees and larger than $(1 + \beta1 + \beta1)$ for 4-year degrees. Substituting .15 and .55 for Pl and Ph, and using the estimates from column 5 of Table 4A, this expression can be solved for l, and the proportion of part-time workers in the low relevance groups, l/Pl, as well as in the high relevance group $(.158 − l)/Ph$, can both be calculated.

Table 7A
Gender differences in estimated relationship between earnings, schooling, and occupational relevance of schooling.

	Schooling = 4-year college degree		Schooling = 4-year college degree	
	(1)	(2)	(3)	(4)
Schooling	0.067 (1.015)	0.036 (0.474)	0.015 (0.623)	−0.010 (0.354)
Relevance	0.145 (1.735)*		0.018 (0.160)	
(Schooling) × (Relevance)	0.193 (1.722)*		0.072 (1.925)*	
Med. Relevance		0.008 (0.120)		−0.147 (1.340)
High Relevance		0.065 (1.038)		−0.069 (0.696)
(Schooling) × * (Med. Relevance)		0.126 (1.344)		0.079 (2.229)**
(Schooling) × (High Relevance)		0.175 (2.016)**	(2.516)**	0.082
Gender Interaction Terms				
(Female) × (Schooling)	−0.083 (0.837)	0.013 (0.107)	−0.019 (0.501)	0.028 (0.604)
(Female) × (Relevance)	0.111 (1.407)		0.073 (0.447)	
(Female) × (Schooling) × (Relevance)	0.103 (0.848)		0.046 (0.837)	
(Female) × (Med. Relevance)		0.060 (0.581)		0.182 (1.074)
(Female) × (High Relevance)		0.129 (1.379)		0.140 (0.906)
(Female) × (Schooling) × (Med. Relevance)		−0.102 (0.693)		−0.062 (1.116)
(Female) × (Schooling) × (High Relevance)		0.015 (0.108)		−0.002 (0.039)
R-squared	0.265	0.263	0.265	0.263

Notes: sample size is 2268 individuals with post-secondary degrees. In the first two columns, "schooling" is an indicator for having a 4-year college degree; in the second two columns, "schooling" is the number of years of post-secondary education. "Med. Relevance" and "High relevance" are 0–1 indicators equal to 1 if degree is "somewhat important" or "important" for "Med. relevance" and "very important" or "extremely important" for "High relevance". The reference level of relevance is "Low relevance". Controls included in all regressions are gender, race, marital and parenting status, union status, an indicator for having multiple post-secondary degrees, and 2-digit occupation and industry controls. Absolute values of t-statistics in parentheses.

* Is significant at 10%.
** Is significant at 5%.

the observed differences in earnings between occupational relevance groups.

To examine this issue further, Tables 7A and 7B report the results of estimating the models with gender interaction effects (Table 7A), and the results of estimating the models after excluding women with children (Table 7B). Both robustness analyses attempt to account for unobserved differences in hours worked by stratifying the sample based on demographic characteristics likely correlated with work hours (for example, in the CPS 1991 sample, women were working 5 h fewer per week on average, and married women with children were working 6 h fewer per week on average, as compared to men, both differences are significant at $p < .001$). We expect that if the relevance earnings premium is driven entirely by unmeasured differences in work hours, the coefficient of the relevance variable and of its interaction term with schooling will become indistinguishable from zero (or at least significantly smaller in magnitude) once demographic groups with high propensity toward part-time work are parsed out. In both tables, however, the difference in earnings between good and bad occupational matches remains statistically significant ($p < 0.10$ in Table 7A and $p < 0.01$ in

Table 7B). The changes in the magnitudes of the coefficients between Tables 5A and 5B and Tables 7A and 7B, although they are in the expected direction, are not statistically significant. For example, the relevance premium for a 4-year degree decreases from 40.8% in Table 5A to 33.8% (.145 + .193) in Table 7A (p-value of the difference is .44), and to 36.9% (.214 + .155) in Table 7B (p-value of the difference is .76). Furthermore, if the earnings premium were a result of the unmeasured variation in work hours, we would expect the return to relevance to be significantly higher for females than for males. However, the difference in the relevance earnings premium between females and males, as estimated by the interaction terms shown in the lower half of Table 7A, although positive, is not significant in any of the models (p-values of the coefficients range between .25 and .51). These results provide further evidence to support the notion that the relevance earnings premium not likely to be fully attributable to unmeasured systematic variation in work hours.

In the end, we found that people who matched their degree to an occupation earn significantly more than people who work in occupations unrelated to their degree field, and that the positive earnings differentials are robust to

Table 7B
Impact of excluding women with children on the estimated relationship between earnings, schooling, and occupational relevance of schooling.

	Schooling = 4-year college degree		Schooling = Years of PSE	
	(1)	(2)	(3)	(4)
Schooling	0.036	0.059	−0.002	−0.003
	(0.689)	(0.961)	(0.097)	(0.128)
Relevance	0.214		0.109	
	(3.959)***		(1.269)	
(Schooling) × (Relevance)	0.155		0.067	
	(2.085)***		(2.379)***	
Med. relevance		0.086		0.022
		(1.637)		(0.251)
High relevance		0.165		0.060
		(3.429)***		(0.772)
(Schooling) × (Med. Relevance)		0.008		0.024
		(0.108)		(1.880)*
(Schooling) × (High Relevance)		0.121		0.061
		(1.807)*		(2.386)**
R-squared	0.266	0.267	0.264	0.264

Notes: women with children are excluded, sample size is 1918 individuals with post-secondary degrees. In the first two columns, "schooling" is an indicator for having a 4-year college degree; in the second two columns, "schooling" is the number of years of post-secondary education. "Relevance" is a continuous variable. "Med. Relevance" and "High relevance" are 0–1 indicators equal to 1 if degree is "somewhat important" or "important" for "Med. relevance" and "very important" or "extremely important" for "High relevance". The reference level of relevance is "Low relevance". Ability/aptitude/motivation controls are included in all regressions. Other controls included in all regressions are gender, race, marital and parenting status, union status, an indicator for having multiple post-secondary degrees, and 2-digit occupation and industry controls. Absolute value of *t*-statistics in parentheses.

  * Is significant at 10%.
 ** Is significant at 5%.
*** Is significant at 1%.

controlling for positive selection on unobservable characteristics and for earnings differences across degree fields. Thus, our findings are consistent with the idea that human capital underutilization can play a role in the association between poor quality of occupational match and low returns to schooling.

5. Conclusion

This paper examines whether the return to schooling varies depending on the extent to which a person's occupation matches the field of their post-secondary training using a new measure of occupational match quality. We find that it does, as people who match their schooling to an occupation earn a higher return to their post-secondary educational investment than people who work in occupations that do not draw on their specialized human capital.

While personal characteristics are, without a doubt, strong predictors of success in the labor market, we were unable to rule out the possibility that factors exogenous to an individual may systematically interfere with a person's transition from the school desk to the workplace. In particular, we found that mismatch between the field of schooling and occupation was strongly associated with low earnings and low returns to schooling, even after we controlled for many of the individual differences that may lead to poor labor market experiences, and for differences in the marketability of degree fields. Although the true extent of the causal link between the quality of occupational match and the return to schooling may not be entirely clear, this study does emphasize the potential importance of policy interventions aimed at improving the efficiency and transparency of labor markets.

This work leaves an important question open for further research. Specifically, we need to develop a better understanding of the reasons why some people with specialized training are employed in occupations in which their skills are underutilized. Efforts to answer this question will not only expand academic knowledge of links between people's educational and labor market experiences; they may also reveal potentially modifiable factors that may someday lay foundation for new public policies.

On a final note, the reader may rightfully point out that adding controls for ability, aptitude, and motivation, regardless of how rich a set of controls is available to the researcher, does not eliminate unobserved individual heterogeneity entirely, and that some residual ability bias in the estimates of the coefficients is likely to still remain. However, given that the additional controls noticeably decrease the size of the overall return to schooling while leaving the return to relevant degrees virtually unchanged, we believe it is unlikely that the observed positive earnings differential associated with better occupational matching is driven entirely by residual ability differences among individuals. At the very least, this study makes a case for the role of occupational matching as a factor in the relationship between schooling and earnings, and is an important first step toward a better understanding of the mechanism of this relationship.

References

Arias, O., Hallock, K. F., & Sosa-Escudero, W. (2001). Individual heterogeneity in the returns to schooling: Instrumental variables quantile regression using twins data. *Empirical Economics*, *26*, 7–40.

Card, D. (1995). Earnings, schooling and ability revisited. In S. Polachek (Ed.), *Research in labor economics, vol. 4* (pp. 23–48). JAI Press.

Card, D. (1999). The causal effect of education on earnings. In O. Ashenfelter, & D. Card (Eds.), *Handbook of Labor Economics, vol. 3A* (pp. 1801–1864). North Holland.

Clogg, C. C., & Shockey, J. W. (1984). Mismatch between occupation and schooling: A prevalence measure, recent trends and demographic analysis. *Demography*, *21*(2), 235–257.

Halaby, C. H. (1994). Overeducation and skill mismatch. *Sociology of education*, *67*(1), 47–59.

Hugerford, T., & Solon, G. (1987). Sheepskin effects in the return to education. *Review of Economics and Statistics*, *69*, 175–177.

Jaeger, D. A., & Page, N. (1996). Degrees matter: New evidence on sheepskin effects in returns to education. *Review of Economics and Statistics*, *78*(4), 974–987.

Manacorda, M., & Petrongolo, B. (1999). Skill mismatch and unemployment in OECD countries. *Economica*, *66*(262), 181–207.

Miller, R. A. (1984). Job matching and occupational choice. *The Journal of Political Economy*, *92*(6), 1086–1120.

Mincer, J. (1974). The human capital earnings function. In *NBER book schooling, experience and earnings*. Columbia University Press., pp. 83–96

Rumberger, R. W. (1987). The impact of surplus schooling on productivity and earnings. *The Journal of Human Resources*, *22*(1), 24–50.

Shockey, J. W. (1989). Overeducation and earnings: A structural approach to differential attainment in the US labor force (1970–1982). *American Sociological Review*, *54*(5), 856–864.

Sicherman, N. (1991). Overeducation in the labor market. *Journal of Labor Economics*, *9*(2), 101–122.

Smith, H. (1986). Overeducation and underemployment: An agnostic view. *Sociology of Education*, *59*, 85–99.

Spence, M. (1973). Job market signaling. *The Quarterly Journal of Economics*, *87*(3), 355–374.

Sullivan, T. A. (1978). *Marginal workers, marginal jobs*. Austin, TX: University of Texas Press.

Vahey, S. P. (2000). The great Canadian training robbery: Evidence of returns to education mismatch. *Economics of Education Review*, *19*(2), 219–227.

Verdugo, R. R., & Verdugo, N. T. (1989). Overeducation and the earnings of Black, Hispanic, and white male workers. *Sociological Perspectives*, *31*(2), 190–212.