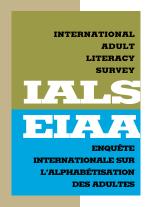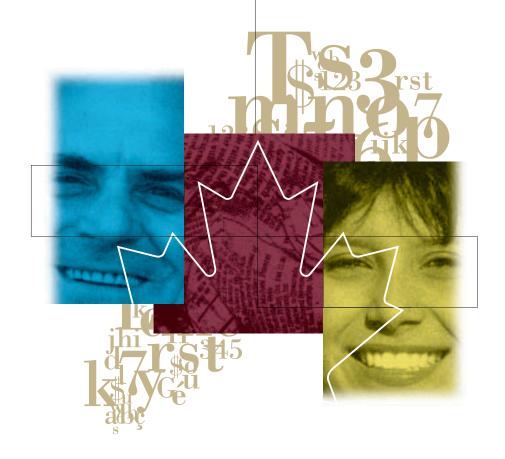# Literacy Skills, Occupational Assignment and the Returns to Over- and Under-Education

Daniel Boothby

Statistics
Canada

Statistique
Canada

Human Resources
Development Canada

Développement des ressources
humaines Canada

Canada

How to obtain more information

Specific inquiries about this product and related statistics or services should be directed to: T. Scott Murray, Statistics Canada, Ottawa, Ontario, K1A 0T6 (telephone: (613) 951-9035).

For information on the wide range of data available from Statistics Canada, you can contact us by calling one of our toll-free numbers. You can also contact us by e-mail or by visiting our Web site.

National inquiries line	**1 800 263-1136**
National telecommunications device for the hearing impaired	**1 800 363-7629**
Depository Services Program inquiries	**1 800 700-1033**
Fax line for Depository Services Program	**1 800 889-9734**
E-mail inquiries	**infostats@statcan.ca**
Web site	**www.statcan.ca**

Ordering and subscription information

This product, Catalogue no. 89-552-XPE, is published (irregularly) as a standard printed publication at a price of CDN $10.00 per issue. The following additional shipping charges apply for delivery outside Canada:

Single issue

United States	CDN $ 6.00
Other countries	CDN $ 10.00

This product is also available in electronic format on the Statistics Canada Internet site as Catalogue no. 89-552-XIE free of charge. To obtain an issue visit our Web site at **www.statcan.ca**, and select Products and Services.

All prices exclude sales taxes.

The printed version of this publication can be ordered by

- Phone (Canada and United States) **1 800 267-6677**
- Fax (Canada and United States) **1 877 287-4369**
- E-mail **order@statcan.ca**
- Mail Statistics Canada
  Dissemination Division
  Circulation Management
  120 Parkdale Avenue
  Ottawa, Ontario K1A 0T6
- And, in person at the Statistics Canada Regional Centre nearest you, or from authorized agents and bookstores.

When notifying us of a change in your address, please provide both old and new addresses.

Standards of service to the public

Statistics Canada is committed to serving its clients in a prompt, reliable and courteous manner and in the official language of their choice. To this end, the Agency has developed standards of service which its employees observe in serving its clients. To obtain a copy of these service standards, please contact Statistics Canada toll free at 1 800 263-1136.

The paper used in this publication meets the minimum requirements of American National Standard for Information Sciences - Permanence of Paper for Printed Library Materials, ANSI Z39.48 - 1984.

∞

International Adult Literacy Survey

Literacy Skills, Occupational Assignment and the Returns to Over- and Under-Education

Daniel Boothby

Human Resources Development Canada

Published by authority of the Minister responsible for Statistics Canada

The data interpretations and policy prescriptions presented in this report are those of the author and do not necessarily reflect those of the granting agencies or reviewers.

Acknowledgements

Note of Appreciation

Canada owes the success of its statistical system to a long-standing partnership between Statistics Canada, the citizens of Canada, its businesses, governments and other institutions. Accurate and timely statistical information could not be produced without their continued cooperation and goodwill.

Table of contents

Abstract

This study uses data from the Canadian panel of the International Adult Literacy Survey to examine the relations between schooling, literacy and occupational assignment and to determine the extent to which returns to over- and under-education are in fact returns to literacy skills.

Two measures of required training time for the job are used, both of which are based on detailed occupation. One is the General Educational Development (GED) level of the occupation; the other is the sum of the GED and Specific Vocational Preparation (SVP) times.

Regression analysis of the required training time of individuals' jobs shows that literacy skills are an important determinant of occupational assignment by skill level, once schooling is taken into account. Skills acquired through on-the-job training may also play an important role in occupational assignment.

The research literature on returns to over-education and under-education examines the relation between workers' skills, as measured by their level of schooling, and the skill requirements of their job. The "typical" findings in this body of research are 1) that over-educated workers (schooling greater than required by their job) earn more than workers in jobs with comparable educational requirements but with the (lower) schooling levels that match these requirements; 2) that over-educated workers earn less than workers with comparable schooling in jobs which require this level of schooling; and 3) that under-educated workers (schooling less than required by their job) earn more than comparably educated workers in jobs which match their schooling, but less than workers in jobs with comparable educational requirements whose schooling matches these requirements.

We find this pattern of returns to over-education and under-education for women and men in our sample using regression analysis of the (log of) earnings of full-time workers. When measures of literacy skill are added to these regressions, the estimated coefficients of both over-education and under-education decrease in absolute value for men and the estimated coefficients of under-education increase for women. When a measure of literacy use at work is added, this variable has a positive coefficient; and there are further decreases in the absolute values of the coefficients of both over- and under-education for both women and men.

We conclude that literacy skills play a significant role in occupational assignment, independent of the role of schooling, that the return to under-education for both women and men is in large part a return to above average literacy skills for their level of schooling, and that for men, the return to over-education is in large part a return to literacy skills which are above average for their jobs. This would seem to indicate that employers are capable of determining their employees' literacy skills by more accurate means than simply depending on the level of schooling as an indicator of literacy skills.

Chapter 1

Introduction and Literature Review

Issues concerning skill—what it is, how it can be measured, how it is acquired, how it is rewarded—are of great importance for policy-oriented labour market research. This paper examines an empirical question, whether the observed returns to "over-education" and "under-education" are in fact returns to literacy skills. As will be suggested below, the answer to this empirical question has implications for our view of skill.

The research reported here draws on several lines of research on skill. As the title suggests, the first of these is research on "over-education", "under-education" and the returns to "over-education" and to "under-education".

The notion of "over-education" is derived from the observation that there are large numbers of persons working in jobs whose educational requirements are less than the educational attainments of the job-holders. This idea has a popular and journalistic version, that large numbers of university graduates are working as taxi drivers.

Among academic writers, David Livingstone (1999) has recently published a book entitled *The Education-Jobs Gap*. He provides a time series of comparable data for Ontario which shows that in a survey conducted at two year intervals, 18-23% of the Ontario work force is in jobs where they are "under-employed" and 22-26% is in jobs in which they are "under-qualified". (Table 2.6, 76) Comparable data from the *General Social Survey* would also seem to indicate that about 20% of the work force is "under-employed". Vahey (2000), using self-assessment data from the *National Survey of Class Structure and Labour Process in Canada*, finds that 30% of male workers and 32% of female workers were over-educated in 1982, while 24% of males and 17% of females were under-educated. He notes that almost all mismatch was of one schooling level, for example, community college graduates working in jobs they assessed as requiring a high school education.

This phenomenon is of public and policy concern because of the perception that "over-education" is a wastage of private and social resources devoted to education. Note that the idea that the 20% or so of "over-educated workers" are in jobs where their education is wasted rests on the following assumptions: that jobs have a level of educational requirements, that this level is measured accurately, that persons with more education cannot perform this job better than at the

level implied by the educational requirements, and that education is an accurate measure of an individual's ability to meet the skill requirements of the job. If one adheres to this view of the labour market, one should be as concerned about "under-education" as about "over-education", since it would indicate the presence of a considerable proportion of the work force whose skill levels do not match those required by the job.

Returns to Over-Education and Under-Education

The results of the research literature on "returns to over-education" pose a challenge to the view that "over-education" is wastage. A typical set of findings in this literature is that:

1. "over-educated" workers earn more than workers in jobs with comparable educational requirements but with (the lower) schooling levels that match these requirements;

2. "over-educated" workers earn less than workers with comparable schooling in jobs which require this level of schooling; and

3. "under-educated" workers earn more than comparably educated workers in jobs which match their schooling, but less than workers in jobs with comparable educational requirements whose schooling matches these requirements.

More formally, following Sicherman (1991) let S be an individual's years of schooling and let R be the years of schooling required in the individual's job. Years of over-education O is defined as O = {S-R when positive, 0 otherwise}. Years of under-education U is defined as U = {R-S when positive, 0 otherwise}. In an earnings regression which includes R, O and U as regressors, call the estimated coefficients of these variables r, o and u respectively. The statements above amount to 1) $r > o > 0$, 2) $u < 0$, and 3) $r + u > 0$ (so that the net return to a year of under-education is positive).

Sicherman (1991) regresses the log of the hourly wage on a series of standard explanatory variables and on R, O and U for a sample of American men drawn from the *Panel Study of Income Dynamics*.

The estimated coefficients are r = .048, o = .039, u = -.017. All of these are statistically significant. The hypothesis that r = o is strongly rejected by a Wald test.

In his introductory essay to a recent special edition of the *Economics of Education Review* on "The economics of over- and under-schooling", Hartog (2000) reviews a series of studies of this type. He concludes (p135) that findings such as Sicherman's are typical and that this specification of the earnings equation is superior when tested against the "Mincer specification (all coefficients equal).'[1]

In one of the articles in this issue, Vahey (2000), using Canadian data for 1981 from the *National Survey of Class Structure and Labour Process,* finds evidence of a positive return to over-education for males with a bachelor's degree, some evidence of a wage penalty for under-educated males in jobs requiring low levels of education and no evidence of returns to over-education or under-education for women.[2] To our knowledge, this is the only study other than ours that provides estimates of the returns to over-education and under-education for Canada.

Over-Education, Under-Education and Literacy

Why should over-educated workers earn any return to their years of schooling beyond the requirements of their job unless this "surplus" schooling contributes to their productivity? Why should under-educated workers earn less than appropriately educated workers in jobs with similar schooling requirements unless the under-educated workers are less productive?

Borghans and de Grip (2000) develop a model in which workers with different levels of education have different productivities in the same job. Workers are specialized to a range of occupations. Their model allows for the possibility of over-education. But even in a situation of

over-education, the additional education is not entirely wasted, as it results in higher productivity on the job and in higher earnings for the over-educated worker than for co-workers with the "right" amount of education.

Their model thus offers an explanation for the observed pattern of returns to over-education. Over-educated workers receive more than their co-workers because they are more productive. They receive less than workers with a similar educational level working in jobs requiring that educational level because they are constrained by their type of training to work in occupations in which education is in relative over-supply. They contrast this with educational upgrading, in which a change in technology increases the demand in an occupation for more educated workers.

Sicherman (1991) offers a view of over-education and under-education as reflecting substitution between different methods of acquiring human capital. He finds support for the hypotheses that on average over-educated workers have less work experience and under-educated workers have more work experience than workers with the appropriate schooling level for their job. He also finds support for the hypothesis that over-educated workers may be working temporarily in jobs in which they can acquire work skills needed for upward occupational mobility. Finally, he finds support for a job-matching view in which the poor job/skills match of over-educated workers makes them more likely to leave their jobs. To the extent that these poor job matches are temporary, the "wastage" due to over-education will be small.

Pryor and Scheffer (1997, 1999) take a very different view of over-education. They seek to explain why college-educated workers are working in "high-school jobs" in the U.S., particularly when a growing gap between the wages of the high-school-educated and the college-educated would seem to indicate a rising relative demand for the college-educated. Their answer is that the rising demand is a demand for cognitive skills, that on average the college-educated have higher levels of cognitive skills, but that the college-educated working in high-school jobs have levels of cognitive skills appropriate to high school graduates. They provide support for this view using data from the *National Adult Literacy Survey* that show lower average levels of literacy skills among college graduates working in high-school jobs than among high school graduates working in college jobs.

Green and Riddell (2001) provide evidence for Canada from the 1994 *International Adult Literacy Survey* (IALS) that literacy skills have an impact on earnings which accounts for about a third of the total effect of education on earnings. Osberg (2000) explores the relation between earnings, schooling and literacy scores using a variety of transformations of the sum of the three literacy scores from the Canadian IALS data. He concludes that whatever the transformation used, a considerable proportion of the return to education is a return to literacy skills, particularly for males. Charette and Meng (1998) use Canadian data from the 1989 *Survey of Literacy Skills Used in Daily Activities* (LSUDA). They find that including literacy and numeracy measures in income equations increases the return to schooling for women and decreases the return to schooling for men.[3]

Boothby (1999) provides evidence that low levels of literacy skill are related to job-education mismatch for Canada. Using data from the 1994 *International Adult Literacy Survey* (IALS) he shows that the predicted probability that a university graduate is "mismatched" decreases strongly as literacy skills increase.[4] He also shows that "occupational mismatch" has a negative effect on earnings for post-secondary graduates.

Can Differences in Literacy Skills Explain the Returns to Over-Education and Under-Education?

Although they do not make this point, the results given by Pryor and Shaffer (1999) provide a possible explanation for the returns to "over-education" and to "under-education". First, they show (Table 3.5, 66) that the average literacy levels of college graduates are higher in occupations with higher educational levels than in occupations with lower educational levels. The same table shows,

however, that within each educational level of jobs, average literacy scores increase with the schooling level of job-holders. Second, they show (Table 5.1, 104-105) for all levels of educational attainment (including university graduates), a strong relation between wages and the educational level of the job. The same table also shows that within each educational tier of jobs, wages increase with increasing educational attainment.

This suggests that the following mechanism might be at work. Workers with low levels of literacy skills for their educational attainment may match to jobs in which the normal educational level is below their educational level. They will be observed to be over-educated in these jobs. They may, however, have higher levels of literacy skills than those workers with the "appropriate" level of education for the job. The higher than average literacy skills (for the educational level of the job) will earn a return in higher than average wages (for the educational level of the job). This will be observed as a return to over-education. It will be lower than the return to required education, because their literacy skills are lower than those of persons with the same level of education working in jobs at the appropriate educational level.

Conversely, under-educated workers will be those with the highest levels of literacy skills in their educational level. They will match to jobs with "educational requirements" above their schooling level, but will have below average literacy skills in these jobs.

The model described above builds on the view of Pryor and Shaffer that both occupational assignment and earnings are related to literacy skills. In principle, it is not restricted to literacy skills. The same type of model of over-education and under-education could be built in terms of communications skills, or any other skill that might be correlated with schooling. In testing the model, we are limited to a consideration of literacy skills, because these are the skills for which direct measures are available.

The model is similar to that of Borghans and de Grip (2000) (discussed earlier), in that there is no fixed level of skill in an occupation. Instead, workers at different levels of skill can perform in the occupation at different productivity levels with resulting differences in earnings.

The empirical portion of this work will thus consist of using the IALS data to determine:

1. the influence of respondents' schooling, literacy skill and work experience on the skills requirements of their jobs,

2. whether the IALS data show the usual pattern of returns to over-education and to under-education, and

3 the extent to which returns to literacy skills account for returns to over-education and to under-education.

As should be clear from the discussion above, this paper is very much concerned with the relation between skill use and returns to skill. Krahn and Lowe (1998) found that a significant proportion of employed workers in the IALS data set have high levels of literacy skill that they do not use at work. Green et al (2000) divided employed persons from the British sample of the IALS by literacy skill levels and by levels of use of literacy skills at work. They found substantial earnings losses for high-literacy-skills workers in jobs with low use of literacy skills. We therefore also will examine the relation between earnings and the use of literacy skills at work.

The plan of the paper is as follows. In Chapter 2, we will describe our data sources, our measure of the educational requirements of the job and the resulting measures of over-education and under-education. In Chapter 3, we present results of estimates of the determinants of the educational requirements of the respondents' job. In Chapter 4, we report the results of estimates of a series of alternative models of returns to education, literacy, over-education, under-education and use of literacy skills at work. Our conclusions follow Chapter 4.

Chapter 2

Data Sources and the Measures of Required Schooling, Over-Education and Under-Education

Data Source

Our primary data source is the master file version of the Canadian panel of the 1994 *International Adult Literacy Survey* (IALS). This data set collected information on a standard set of schooling, demographic and labour market variables from a representative national sample of Canadians. Tests of literacy skills were also administered to respondents. Three types of literacy skills were measured: document literacy, prose literacy and quantitative literacy. The Canadian data from the IALS are documented in Statistics Canada (undated). The most thorough description and documentation for the IALS is Murray et al (1998).

We restricted our sample to non-students as the relation between educational attainment and the educational level required by the job is likely to be very different for students and non-students. We restricted the sample to native-born Canadians, since the results of Green and Riddell (2001) and of Kapsalis (2000) indicate that the relations between schooling, literacy and earnings may differ significantly between immigrants and the native-born. We required that individuals in the sample be employed full-time, as preliminary estimates showed that the relations between schooling, occupational assignment and earnings are very different for part-time workers. We would attribute this to a heavy concentration of part-time workers in a small number of occupations with low schooling requirements, regardless of the educational attainment of these workers. We also required that the file provide a valid occupational code for individuals in the sample.[5]

Measures of Required Schooling, Over-Education and Under-Education

The data on the master file include detailed occupational codes for the jobs held by employed persons. The occupational coding system used is the 1980 Standard Occupational Classification (1980 SOC). (Statistics Canada, 1981) This allows us to merge estimates of required training

times for detailed (1980 SOC unit group) occupations into our data set. These training time estimates are based on training requirements for highly detailed occupations given in the Canadian Classification and Dictionary of Occupations (CCDO) (Manpower and Immigration, 1971 and 1973).[6]

The CCDO provides two types of training time requirements for highly detailed (7-digit) occupations. These are General Educational Development (GED) and Specific Vocational Preparation (SVP). Each of these two training times was translated into years at the level of highly detailed occupations, then averaged across the highly detailed occupations within the unit group to obtain training time estimates for the unit group occupations.[7] Measures of this type have been widely used in studies that seek to identify changes in the skill requirements of the economy over time. (For Canada, for example, in Hunter and Manley (1986), Myles (1988) and Boyd (1990)).

Years of Schooling and Years of Required Training Time

We used two alternative measures of required training time in detailed occupations. The first is the GED level of the occupation, translated into years. The second is the sum of GED (in years) and SVP in years for the occupation. Table 1 gives for women and men employed full-time their average number of years of schooling and the average number of years of schooling required for their job by highest completed level of schooling. Average levels of both measures of required training time (GED and GED+SVP) are reported in Table 1.

Each of these measures has drawbacks. The GED measure concerns only general education, and does not account for any career-specific training received in educational institutions. This measure is thus likely to underestimate the schooling requirements for occupations, especially those with high levels of specialized training in educational institutions. SVP accounts for time spent training for specific occupations in educational institutions. The difficulty with this measure is that it also takes into account time spent in other forms of occupationally specific training, for example, training on the job. Consequently, the sum of GED and SVP may overstate the schooling requirements of many occupations, in particular, those occupations in which many skills are acquired through on-the-job training, apprenticeship, and so on.

Table 1 **Mean Levels of Years of Schooling, GED and GED+SVP by Highest Completed Level of Schooling and Gender**

Highest Completed Level of Schooling	Men			Women		
	Years of Schooling	GED	GED+SVP	Years of Schooling	GED	GED+SVP
All	13.1	11.6	14.2	12.8	11.7	13.7
Less than Primary	4.5	8.5	9.2	6.1	9.2	9.5
Primary	8.4	9.8	11.5	8.1	9.5	10.1
Some Secondary	9.9	10.1	11.6	10.1	10.5	12.1
Secondary	12.4	11.1	13.5	12.1	11.5	13.4
Non-University Post-Secondary	14.4	11.9	14.8	14.2	12.0	14.0
University Undergraduate	17.2	13.6	17.1	17.5	13.8	16.8
Post-Graduate	18.8	15.1	20.8	17.9	14.4	17.1
	n = 1,046	n = 1,049	n = 1,049	n = 914	n = 915	n = 915

The average number of completed years of schooling by highest completed level of schooling in Table 1 is much as expected and is very similar for women and men. It is worth noting that university graduates without a post-graduate qualification report over 17 years of completed schooling on average. This will mean that university graduates are on average over-educated by the GED measure, since the maximum required level of schooling with this measure is 17 years.

Average required years of training time increase for both measures as the highest level of completed schooling rises. The levels of required training time are very similar for women and men with the same highest completed level of schooling. Not only does the average level of each measure of training time rise as the highest completed level of schooling increases, but the gap between the two measures (SVP time) increases.

Women and men who have completed some secondary education work in jobs that require on average about ten years of GED and about 11.5 years of GED plus SVP (thus about 1.5 years of SVP). The difference between the two measures is more than two years for high school graduates and non-university post-secondary graduates and more than three years for university graduates. Thus with rising educational levels, required training time as measured by GED plus SVP increases much more rapidly than required schooling as measured by GED.

Over-Education or Under-Education?

These differences in the two measures of required schooling lead to important differences in the resulting measures of over-education. The average number of years of over-education or under-education for a table cell can be computed by subtracting the average number of years of attained schooling from the average required years of schooling. Using the GED measure of required schooling, women and men graduates at every level of post-secondary schooling have on average more than two years of over-education. Using the GED+SVP measure, there is a fairly close match between the average years of schooling of post-secondary graduates and the average years of training required by their jobs.

High school graduates are over-educated on average by the GED measure and under-educated on average by the GED+SVP measure. The average years of schooling of persons with some secondary education about matches their average required years of schooling as measured by GED, but is about two years less than the required training time as measured by GED+SVP. (Some or all of the difference between years of schooling and years of GED+SVP for these two groups may be accounted for by specific vocational preparation received outside of academic institutions).

The two measures of educational requirements thus give conflicting results as to the extent of over-education and under-education in the work force. The GED measure implies that women have a year of over-education and men have a year and a half of over-education on average. The GED+SVP measure implies that both men and women have a year of under-education on average. The measures agree in showing that over-education rises (under-education falls) as the highest level of completed schooling increases.

These two measures share certain difficulties. Both describe an average level of training time requirements. Differences in training time requirements between jobs within a detailed occupation are not captured. GED and SVP are based on analyses of occupations first published in the early 1970's. They do not capture any increase in the training time requirements of occupations.

Reasons like these lead Hartog (2000, 132-133) to conclude that it is preferable to use worker assessment of the educational requirements of the job in measuring over-education and under-education. This type of measure also has its drawbacks. It may reflect employer hiring practices based on the availability of persons with a given level of education, rather than genuine requirements of the job. It may also create spurious variance in job requirements.[8]

In any event, the IALS did not ask employed individuals to evaluate the educational requirements of their job, so that a measure of required schooling based on worker assessment is not available. As Hartog states (2000, 133): "Can we prefer one measure over the other? Usually the choice is dictated by data availability...". Our use of training time requirements based on occupation is dictated by the fact that these are the only available measures of the educational requirements of the job.

Since we have no strong reasons to prefer one or the other of our two job requirements measures, we have used both in what follows. As stated, one of these measures implies that the Canadian workforce in aggregate is significantly over-educated for the jobs in which they work; the other implies significant under-education in aggregate. This divergence serves to emphasize the inexact character of the measures of required schooling we use. In light of the approximate character of these measures, it seems to us useful to report results based on both.

Chapter 3

Determinants of the match to required schooling

Literacy and the match to required schooling

This section examines the relation between the observed characteristics of individuals and the required level of schooling of their job. Individuals choose a job among the jobs available to them. Employers are assumed to employ only individuals who have the skills required by the job, to prefer more-skilled to less-skilled individuals at the same wage, and to be willing to pay a higher wage to more-skilled individuals than to less-skilled individuals for a given job. The implications are that jobs with higher skill requirements will employ more-skilled workers and that workers at different skill levels may work in the same job at differing wage rates. The assumptions underlying this description of employer's behaviour are similar to the concept of education-productivity profiles for occupations discussed in Borghans and de Grip (2000, 6-10).[9]

Schooling is assumed to produce skills, among them literacy skills. Persons with the same amount of schooling may have differing levels of various skills produced by schooling, including literacy skills. Other types of human capital investment, including on-the-job training, also produce skills. Some of the skills produced by on-the-job training may not be available through schooling. Jobs with higher levels of schooling requirements may also require higher levels of skill produced by other types of human capital investment. (The correlation between GED and SVP may be due in part to skill complementarities of this type). Sicherman (1991) discusses many of these points and their implications for modeling job-matching, over-education and under-education.

Among the jobs available to them, that is, those for which they meet the employer's skill requirements, individuals are assumed to choose the job with the highest pay level. Pryor and Schaffer (1997, 1999) find that workers at a given schooling level with below average literacy skills for their schooling are likely to be employed in jobs with educational requirements below their educational level. Boothby (1999) reports similar findings for workers with a post-secondary education in Canada.

Either of two reasons might explain the concentration of workers with below-average literacy skills for their schooling level in jobs with educational requirements below their schooling level. First, workers at the lower end of the literacy distribution for a given level of schooling may not

meet the minimal skill requirements of employers for jobs whose schooling requirements match these workers' level of schooling. Second, the skill-wage profiles in various occupations might be such that these workers can earn more in occupations with schooling requirements below their schooling levels. This would be the case if the return to over-education in occupations where these workers are over-educated more than offsets the wage penalty to below-average literacy for these workers in occupations with schooling requirements that match their schooling level.

In either case, one would observe that "over-educated" workers are those with below average literacy skill levels for their schooling. If "over-educated" workers have above average literacy skills for the occupations in which they work, the "return to over-education" might be a return to these skills, at least in part.

Similarly, under-educated workers would be those with above average levels of literacy skill for their schooling. These higher skill levels would have allowed them access to occupations with higher required schooling levels. The observed pattern of returns to under-education would be due to lower literacy skill levels of under-educated workers than of workers whose schooling level matches the requirements of the occupation.

The implications of this view for the relation between workers' characteristics and the schooling requirements of their job are straightforward. First, the required schooling of the job is a measure of the level of skills required by the job, including literacy skills. Some of these skills may be complementary skills which are acquired through other forms of human capital investment than schooling. Second, workers with more schooling and with more on-the-job training are more likely to be found in jobs with higher skill levels (measured by higher schooling requirements).

Third, adding a measure of literacy skills to the model predicting the skill level of the job should have two effects. Persons with higher literacy skills should work in jobs with higher educational requirements. Adding a direct measure of literacy skills should reduce the effect of schooling on the skill level of the job, since part of the effect of schooling is due to its effect on literacy skills.[10]

Estimates of the determinants of required schooling

Table 2 gives the results of regression estimates of the determinants of the level of educational requirements of the jobs held by women and men employed full-time. Results are reported for both measures of schooling requirements for women and men separately.

Several aspects of the estimated effects of schooling and work experience are worth noting.[11] First, for both women and men, the coefficients of these variables are much larger for the GED+SVP measure than for the GED measure. Second, for both measures, the coefficients of schooling and work experience are much larger for men than for women.

If the coefficient of schooling in the required education regressions were equal to one, a one year increase in schooling would imply a one year increase in required schooling, leaving the amount of over-education (or under-education) unchanged. All of the coefficients of schooling in these regressions are less than one. Thus on average, over-education increases (under-education decreases) as the level of schooling rises. The rise in over-education with increasing schooling is weakest in the regression for the GED+SVP measure for men (coefficient of schooling = .840) and strongest in the regression for the GED measure for women (coefficient of schooling = .390).

Table 2   **Baseline Regressions for Determinants of Schooling Requirements of Occupation**

	Men (n = 925)				Women (n = 836)			
	GED		GED+SVP		GED		GED+SVP	
Constant	*4.415*	*(.451)*	**2.001**	(.868)	*6.809*	*(.435)*	*6.372*	*(.813)*
Years of school	*.500*	*(.024)*	*.840*	*(.047)*	*.390*	*(.026)*	*.590*	*(.049)*
Experience	*.052*	*(.007)*	*.094*	*(.014)*	*.026*	*(.007)*	*.052*	*(.139)*
Atlantic	-.084	(.267)	-.063	(.514)	**-.468**	**(.241)**	-.651	(.452)
Quebec	.039	(.185)	-.053	(.356)	*-.875*	*(.176)*	*-1.39*	*(.330)*
Ontario	—	—	—	—	—	—	—	—
Prairies	.050	(.213)	.178	(.410)	*-.527*	*(.202)*	-.507	(.379)
British Columbia	*-.654*	*(.266)*	**-1.07**	**(.512)**	.047	(.220)	.591	(.412)
Large Urban (>500,000)	—	—	—	—	—	—	—	—
Small Urban (<500,000)	*-.559*	*(.166)*	*-.895*	*(.319)*	*-.464*	*(.157)*	*-.954*	*(.294)*
Rural	*-.756*	*(.213)*	**-1.39**	**(.410)**	*-.676*	*(.185)*	*-1.19*	*(.347)*
Never-married	—	—	—	—	—	—	—	—
Ever-married	.047	(.196)	.109	(.376)	.183	(.183)	.014	(.343)
Widowed, divorced, separated	.466	(.332)	.963	(.638)	.284	(.206)	-.625	(.386)
r² (adjusted for d.o.f.)	.36		.30		.27		.20	

Estimated standard errors in parentheses.

Bold-face and italic: significant at 1%. Bold-face only: significant at 5%.

Higher levels of work experience are associated with higher educational requirements of the job. This may occur because work experience is a partial substitute for education or because certain jobs with high levels of educational requirements have complementary work experience requirements (see Sicherman, 1991). The coefficients of schooling and work experience indicate that it takes nine or more years of work experience to have the same effect as a year of schooling on the educational requirements of the job.

Work experience is associated with a larger rise in the educational requirements of the job for men than for women. One might speculate that this is because the approximation used for work experience (age - years of schooling - 6) overestimates work experience for women, thus biasing the coefficient downwards.[12]

In addition to schooling and work experience, a series of geographic and marital status indicators are included in the regressions reported in Table 2. The marital status variables have no statistically significant effects. Living in a small urban area (less than 500,000 population) or in a rural area has negative, statistically significant effects on the educational requirements of the job. This is perhaps due to differences in the composition of employment in these areas as compared to large urban areas.

Regional effects are very different for women and men.[13] For men, the only statistically significant effect is a negative effect of residence in British Columbia (relative to residence in Ontario). For women, residence in Quebec has a negative, statistically significant effect in both sets of estimates and residence in the Atlantic region or the Prairie region has a negative, statistically significant effect in the estimates for the GED measure.

Table 3 shows the effect on the years of schooling coefficients of adding measures of literacy skills to the regressions reported in Table 2. Two measures of literacy skills are used: the average of the three "first plausible values" for prose literacy, document literacy and quantitative literacy (labelled literacy score) and a measure based on the percentage of correct answers to the test items

in the IALS.[14] This second measure is included because the literacy scores reported in the IALS are in part conditioned on the respondent's level of schooling. The use of a measure based directly on the test items is a means of identifying the possible bias associated with using literacy scores that are conditioned on schooling. (These issues are examined in more detail in Appendix 3).[15]

Table 3 — Effects of Years of Schooling and Literacy Skills on Schooling Requirements of Occupation

	GED						GED+SVP					
Men (n = 925)												
Years of school	*.500*	*(.024)*	*.389*	*(.028)*	*.448*	*(.028)*	*.840*	*(.047)*	*.645*	*(.053)*	*.754*	*(.052)*
Literacy Score	—		*.014*	*(.002)*	—		—		*.025*	*(.004)*	—	
Proportion Correct	—		—		*2.125*	*(.465)*	—		—		*3.537*	*(.897)*
r² (corrected for d.o.f.)	.36		.40		.37		.30		.34		.31	
Women (n = 836)												
Years of school	*.390*	*(.026)*	*.301*	*(.029)*	*.366*	*(.028)*	*.590*	*(.049)*	*.431*	*(.055)*	*.563*	*(.053)*
Literacy Score	—		*.011*	*(.002)*	—		—		*.020*	*(.003)*	—	
Proportion Correct	—		—		.972	(.402)	—		—		1.085	(.754)
r² (corrected for d.o.f.)	.27		.30		.27		.20		.23		.20	

Estimated standard errors in parentheses.

Bold-face and italic: significant at 1%.

In every regression reported in Table 3 in which a measure of literacy skills is included as a regressor, higher literacy skills are associated with working in an occupation with higher schooling requirements. When literacy skills are included, the coefficient of schooling decreases. This suggests that some of the apparent effect of schooling on occupational assignment is in fact an effect of literacy skills on occupational assignment. (To the extent that literacy skills are produced by schooling, the effect of literacy skills is an indirect effect of schooling).

It is also true that in every case the decrease in the coefficient of schooling is greater when the literacy score is used as the measure of literacy than when the percentage of correct answers is used. This might be because the conditioning of literacy scores on schooling results in a downward bias in the coefficient of schooling. It also might be because the regression using the literacy score better measures the effects of schooling (once literacy is taken into account) than does the regression using the percentage correct.

The results reported in Table 2 confirm the hypothesis that skills acquired through schooling and skills acquired through on-the-job training both tend to increase the skill level of jobs to which individuals match, as measured by required training time. The results in Table 3 confirm the hypotheses that literacy skills influence the skill levels of individuals' job and that part of the estimated effects of schooling in the regressions of Table 2 is in fact an effect of literacy skills. All of these effects are greater for men than for women.

Chapter 4

Over-Education, Under-Education, Literacy, Literacy Use and Earnings

Regression results

We turn now to the estimation of returns to over-education and under-education and of the relation of literacy skills to these returns. Table 4 gives baseline estimates of earnings regressions for women and men, prior to the inclusion of any measures of over-education, under-education, literacy skills and the use of literacy at work. The dependent variable is the log of annual earnings.

In our sample from the IALS, there are statistically significant effects of both years of schooling and the square of years of schooling. The coefficient of years of schooling is positive; the coefficient of the square of years of schooling is negative. This pattern of declining returns to schooling differs from the results summarized by Hartog (2000,135). (Issues of functional form are discussed in Appendix 2). The patterns of returns to schooling shown in Table 4 imply that the return to an additional year of schooling is positive up to twenty-one years of schooling for both women and men and is greater for women than for men up to approximately eighteen years of schooling.

The estimated marginal returns to work experience are higher for women than for men, up to thirteen years of work experience.[16] The approximation for work experience used (age - years of school - 6) is likely to over-estimate women's work experience, and thus to result in under-estimation of women's returns to work experience. If a more exact measure of work experience were available for the sample, the gap between women's and men's returns to work experience might well be larger than our estimates.

Patterns of regional variation in earnings are similar for women and men, except that men in British Columbia earn significantly less than men in Ontario, while women in this province do not earn significantly less than women in Ontario. Both men and women working in rural areas earn significantly less than those working in urban areas. Ever-married men earn significantly more than those who have never married, but the current marital status of ever-married men has no significant association with earnings. Marital status variables have no significant associations with earnings for women (it should be recalled that this is a sample of full-time workers).

Table 4   **Baseline Log of Annual Earnings Regressions**

	Men (n= 774)		Women (n= 684)	
Constant	*7.770*	*(.374)*	*6.773*	*(0.455)*
Years of school	*.189*	*(.048)*	*.302*	*(.063)*
Years of school^2	*-.004*	*(.002)*	*-.007*	*(.002)*
Experience	*.043*	*(.009)*	*.062*	*(.008)*
Experience^2	*-.0005*	*(.0002)*	*-.001*	*(.0002)*
Atlantic	*-.372*	*(.088)*	*-.416*	*(.086)*
Quebec	*-.170*	*(.061)*	*-.148*	*(.064)*
Ontario	—	—	—	—
Prairies	-.090	(.071)	*-.157*	*(.074)*
British Columbia	*-.410*	*(.087)*	.094	(.079)
Large Urban (>500,000)	—	—	—	
Small Urban (<500,000)	-.056	(.054)	-.017	(.056)
Rural	*-.158*	*(.073)*	*-.174*	*(.065)*
Never-married	—	—	—	—
Ever-married	*.550*	*(.075)*	-.034	(.071)
Widowed, divorced, separated	-.090	(.112)	-.043	(.072)
r² (adjusted for d.o.f.)	.28		.29	

Estimated standard errors in parentheses.

Bold-face and italic: significant at 1%. Bold-face only: significant at 5%.

Table 5 shows the effects of adding measures of over-education, under-education, literacy skills and literacy use at work to the earnings regressions of Table 4. Aside from the coefficients of these measures, only the estimated coefficients of the years of schooling and its square are shown in Table 5.[17]

The functional form of the regressions reported in Table 5 that include measures of over-education and under-education differs from that used by Sicherman (1991) which we discussed above. The model estimated by Sicherman can be written:

1. $\ln Y = c + rR + oO + uU + X'B + e$

where Y is earnings, R is required education in the individual's occupation, O is years of over-education, U is years of under-education, r, o and u are their respective coefficients, c is the regression constants, X is a vector of other explanatory variables, B a vector of their coefficients and e the error term (individual subscripts are omitted for compactness). Let S be years of schooling. By definition, O = {S-R when positive, 0 otherwise} and U = {R-S when positive, 0 otherwise}. This can be rewritten as

2. $\ln Y = c + rS + o'O + u'U + X'B + e$

where o´ := o-r and u´ := r + u.

Recall that in equation 1 the predicted pattern of signs is r > o > 0, u < 0, r + u > 0.

This yields the following predictions for equation 2: r > 0, o´ < 0, r + o´ > 0, r > u´ > 0. The null hypothesis for equation 1 that Hartog (2000, 135) refers to as the Mincer specification: H_0: r = o = -u is equivalent to the null hypothesis for equation 2: H_0': o´ = 0, u´ = 0.

The coefficient o of O in equation 1 is usually referred to as the return to over-education. We will call the coefficient o´ of O in equation 2 the penalty to over-education, as it is the amount by which the log of earnings is decreased by a year of over-education in equation 2. The coefficient u of U in equation 1 is the penalty to under-education. We will call the coefficient u´ of U the return to under-education, as it is the amount by which the log of earnings is increased by each year of under-education in equation 2.[18]

We preferred a functional form that includes years of schooling explicitly because of the non-linearity of the return to schooling in our sample. The functional form of the regressions for returns to over-education and under-education reported in Table 5 is:

3.　　$\ln Y = c + rS + sS^2 + o'O + u'U + X'B + e$

The vector X of other characteristics includes all of the variables of Table 4 (except the terms in years of schooling). In some of the regressions, it also includes the literacy score or the literacy score and a measure of literacy use.

Table 5	Effects on Log of Annual Earnings of Schooling, Over- and Under-Education, Literacy Skills and Literacy Use									
	(1)		(2)		(3)		(4)		(5)	
Men (n = 774)										
Years of school	*.189*	*(.048)*	*.213*	*(.049)*	*.205*	*(.047)*	*.134*	*(.049)*	*.110*	*(.048)*
Years of school²	*-.004*	*(.002)*	-.004	(.002)	*-.005*	*(.002)*	-.003	(.002)	-.003	(.002)
Years of over-education (GED)	—	—	*-.070*	*(.014)*	—	—	—	—	—	—
Years of under-education (GED)	—	—	*.070*	*(.025)*	—	—	—	—	—	—
Years of over-education (GED+SVP)	—	—	—	—	*-.055*	*(.013)*	—	—	—	—
Years of under-education (GED+SVP)	—	—	—	—	*.030*	*(.008)*	—	—	—	—
Literacy Score	—	—	—	—	—	—	*.003*	*(.001)*	—	—
Proportion correct	—	—	—	—	—	—	—	—	*.989*	*(.157)*
Literacy Use at Work	—	—	—	—	—	—	—	—	—	—
r² (corrected for d.o.f.)	.28		.32		.32		.30		.31	
Women (n = 684)										
Years of school	*.302*	*(.063)*	*.343*	*(.062)*	*.317*	*(.059)*	*.217*	*(.064)*	*.251*	*(.064)*
Years of school²	*-.007*	*(.002)*	*-.006*	*(.002)*	*-.006*	*(.002)*	*-.005*	*(.002)*	*-.006*	*(.002)*
Years of over-education (GED)	—	—	*-.125*	*(.017)*	—	—	—	—	—	—
Years of under-education (GED)	—	—	*.100*	*(.026)*	—	—	—	—	—	—
Years of over-education (GED+SVP)	—	—	—	—	*-.115*	*(.017)*	—	—	—	—
Years of under-education (GED+SVP)	—	—	—	—	*.035*	*(.009)*	—	—	—	—
Literacy Score	—	—	—	—	—	—	*.004*	*(.001)*	—	—
Proportion correct	—	—	—	—	—	—	—	—	*.508*	*(.146)*
Literacy Use at Work	—	—	—	—	—	—	—	—	—	—
r² (corrected for d.o.f.)	.29		.38		.38		.32		.30	

Estimated standard errors in parentheses.

Bold-face and italic: significant at 1%. Bold-face only: significant at 5%. Italic only: significant at 10%.

Table 5   **Effects on Log of Annual Earnings of Schooling, Over- and Under-Education, Literacy Skills and Literacy Use (concluded)**

	(6)		(7)		(8)		(9)	
Men (n = 774)								
Years of school	*.172*	*(.050)*	*.166*	*(.048)*	*.144*	*(.051)*	*.147*	*(.049)*
Years of school²	-.003	(.002)	**-.004**	**(.002)**	-.003	(.002)	**-.003**	**(.002)**
Years of over-education (GED)	*-.061*	*(.014)*	—	—	*-.056*	*(.015)*	—	—
Years of under-education (GED)	**.060**	**(.025)**	—	—	*.047*	*(.025)*	—	—
Years of over-education (GED+SVP)	—	—	*-.049*	*(.013)*	—	—	*-.048*	*(.013)*
Years of under-education (GED+SVP)	—	—	**.026**	**(.008)**	—	—	**.021**	**(.008)**
Literacy Score	*.002*	*(.001)*	*.002*	*(.001)*	*.002*	*(.001)*	*.002*	*(.001)*
Proportion correct	—	—	—	—	—	—	—	—
Literacy Use at Work	—	—	—	—	**.054**	**(.025)**	**.049**	**(.025)**
r² (corrected for d.o.f.)	.32		.33		.33		.33	
Women (n = 684)								
Years of school	*.265*	*(.064)*	*.255*	*(.060)*	*.233*	*(.063)*	*.238*	*(.059)*
Years of school²	**-.004**	**(.002)**	*-.005*	*(.002)*	**-.004**	**(.002)**	*-.005*	*(.002)*
Years of over-education (GED)	*-.123*	*(.017)*	—	—	*-.112*	*(.017)*	—	—
Years of under-education (GED)	**.075**	**(.026)**	—	—	*.047*	*(.026)*	—	—
Years of over-education (GED+SVP)	—	—	*-.114*	*(.016)*	—	—	*-.110*	*(.016)*
Years of under-education (GED+SVP)	—	—	**.026**	**(.009)**	—	—	.013	(.009)
Literacy Score	*.003*	*(.001)*	*.003*	*(.001)*	*.003*	*(.001)*	*.003*	*(.001)*
Proportion correct	—	—	—	—	—	—	—	—
Literacy Use at Work	—	—	—	—	*.134*	*(.025)*	*.138*	*(.025)*
r² (corrected for d.o.f.)	.39		.39		.42		.42	

Estimated standard errors in parentheses.

Bold-face and italic: significant at 1%. Bold-face only: significant at 5%. Italic only: significant at 10%

The first column of Table 5 reproduces the coefficients of years of schooling and years of schooling squared in Table 4. The second and third columns show the effects of adding measures of over-education and under-education for the GED-based measure and the GED+SVP-based measure respectively.

For both women and men, adding measures of over-education and under-education increases the return to a year of schooling significantly. The return to an additional year of schooling which is required by the job is much higher than the return to an additional year of schooling which is not required by the job. (This latter return is the measured return to a year of schooling minus the penalty to over-education.) For both men and women, the return to schooling rises less relative to its level in the baseline regression (column 1) when the GED+SVP-based measure is used (column 3) than when the GED-based measure is used (column 2).

The earnings penalty to over-education is bigger for women than for men and is somewhat larger for the GED-based measure than for the GED+SVP-based measure. The penalty to a year of over-education matches the return to a year of schooling somewhere between fifteen and eighteen years of schooling (depending on the estimate). Up to this point, a year of additional schooling which results in a year of over-education nonetheless raises earnings on average.

The returns to under-education are positive and statistically significant. In other words, there is an earnings gain to individuals who work in an occupation where the educational requirements are greater than their level of education. The earnings gain associated with a year of under-education is greater for women than for men. The return to under-education is greater when the GED-based measure is used than when the GED+SVP-based measure is used.

Column 4 shows the effect of adding the literacy score measure to the basic earnings regression of column 1. As in the estimates of Green and Riddell (2001), the returns to schooling decrease substantially when the literacy score measure is included.

Column 5 uses the measure of literacy skills based on the percentage of right answers, rather than the literacy score measure. As in Table 3, the impact of the percentage of right answers measure on the return to schooling is less than the impact of the literacy score measure. Here as in Table 3, we do not know whether this difference is the result of the conditioning of the literacy score on schooling. Throughout the rest of Table 5 we use the literacy score measure. The regression results in columns 6-9 change very little when the percentage of right answers measure is used.

Columns 6 and 7 include the measures of over-education and under-education, as well as the literacy score measure. Accounting for the effects of over-education and of under-education in columns 6 and 7 results in a substantial increase in the return to schooling over that shown in column 4.

For men, the coefficients of over-education and under-education decrease (in absolute value) from columns 2 and 3 to columns 6 and 7. For women, the coefficient of under-education decreases from columns 2 and 3, but the coefficient of over-education changes very little. For both women and men, the coefficient of the literacy score measure is lower in columns 6 and 7 than in column 4. It should be recalled that the literacy score variable has a significant effect on the educational requirements of the job, independent of schooling. The coefficient of the literacy score in column 4 thus may include indirect effects of literacy on earnings, operating through occupational assignment, as well as direct effects.

Columns 8 and 9 add a measure of literacy use at work to the models of columns 6 and 7. This measure is based on the answers to six items concerning reading at work and four items concerning writing at work.[19] Literacy use has positive, significant effects on earnings for both women and men; these effects are much greater for women.[20] The return to schooling decreases somewhat for men when the literacy use variable is added. There is a small decrease in the return to schooling for women when the literacy use variable is added. The coefficients of the literacy score measure are essentially unchanged from columns 6 and 7 to columns 8 and 9.

The penalties to over-education in columns 8 and 9 are slightly lower than those shown in columns 6 and 7 respectively. The return to under-education decreases when the literacy use variable is added to the model. In column 9 for women, the coefficient of under-education is not statistically significant. This is the only instance in Table 5 in which the coefficient of an over-education measure or an under-education measure is not statistically significant.

Earnings, Literacy, Schooling, and Job Match

Table 5 explores the relation between schooling, job match, literacy and earnings. Two principal questions are addressed: first, whether the measures of over-education and under-education used show the patterns of return to over-education and under-education found by other researchers; second, whether some part of the returns to over-education and under-education (if found) are in fact returns to literacy skills.

Previous research on returns to over-education and under-education (as summarized by Hartog, 2000) finds a positive return to each year of over-education. This return is less than the return to years of schooling which are required by the job. Workers who are under-educated earn a return to working in jobs where the job requirements exceed their schooling, but this return is less than the return to their 'missing years' of required schooling. In other words, a year of over-education earns a positive return but is penalized relative to a year of required schooling; a year of under-education earns a positive return but is penalized relative to the "missing year" of required schooling.

The results reported in Table 5 follow this pattern for both of our measures of over-education and under-education.[21] The penalty to a year of over-education is slightly higher for the GED-based measure than for the GED+SVP-based measure; the return to under-education is higher in the GED-based measure than in the GED+SVP-based measure. (Columns 2 and 3).

The over-education and under-education measures account for the effects of the match between schooling and job requirements. When the effects of job match are taken into account, the estimated return to a year of schooling should be interpreted as the return to a year of schooling which is required by the individual's job. As shown in Table 5, this return is higher than the return to schooling when the job match is not taken into account. (Columns 2 and 3 compared to Column 1).

Adding literacy skills measures to the models of Table 5 lowers the return to schooling. (Columns 4 and 5 compared to Column 1). The explanation of this decrease in the return to schooling offered by Green and Riddell (2001) is that one component of the return to schooling is a return to literacy skills produced by schooling.

Models that include the literacy score measure also show somewhat lower penalties to over-education and returns to under-education for men and returns to under-education for women than models that do not account for literacy skills. (Columns 6 and 7 compared to Columns 2 and 3). It would seem likely, then, that part of the return to under-education is a return to above average literacy skills for the individual's level of schooling. For men, part of the return to over-education may be a return to literacy skills which are above average for men in jobs with the same schooling requirements .

These models also show lower returns to literacy skills than models which include the literacy score measure, but do not include measures of over-education and under-education. (Columns 6 and 7 compared to Columns 4 and 5). We would interpret this to mean that part of the effect of literacy skills on earnings operates through the job match. Tables 2 and 3 show that the level of required schooling time in the job is influenced by literacy skills. The path by which literacy skills affect earnings via the job match thus would be that persons with lower levels of literacy skills obtain jobs with lower required schooling and that earnings are lower in these jobs.

If one is to interpret the earnings return to literacy as a return to skill, it seems plausible that this return will depend on use of the skill. When a measure of literacy use at work is added to the models that include measures of literacy skills, of over-education and of under-education, there is a decrease in the return to schooling and in the return to under-education for both men and women and a decrease in the penalty to over-education for women. Literacy use at work earns a positive return. (Columns 8 and 9 compared to Columns 6 and 7).

Our interpretation of these results is that literacy use at work measures variations in skill requirements of individual jobs that are not captured by our occupation-level measurements of the schooling requirements of the job. Other interpretations are possible. DiNardo and Pischke (1997) find that sitting down at work has a positive effect on earnings. This result serves to emphasize the need for caution in intepreting an association between a work activity and higher earnings as meaning that the ability to carry out the activity is being rewarded by the higher earnings.

Conclusions

Before turning to more specific conclusions, we begin with a general comment on the differences in our results for women and for men. The qualitative results are similar for women and men. There are, however, often large differences between the coefficients of certain of our key variables for women and men. An example is the differences between women and men in the coefficient of schooling in the regressions for the skill level of the occupation. In our view, these differences are in large part due to continuing differences in the occupational distributions of women and men. These differences lead us to report all of the results in the preceding sections by gender.

Our first set of conclusions concerns the determinants of the skill levels of workers' occupations, as measured by the educational requirements of the occupation. We have provided evidence that literacy skills are an important determinant of occupational assignment by skill level, once the role of schooling is taken into account. This agrees with the results of Pryor and Schaffer (1997, 1999) and of Boothby (1999). We have also provided evidence that a significant part of the relation between schooling and occupational assignment (when literacy skills are omitted) is an effect of literacy skills (which are closely related to schooling). Finally, we have provided evidence that skills acquired through on-the-job training also may be a significant determinant of occupational assignment. This agrees with Sicherman's (1991) conclusion that over-education and under-education may result from the effect of skills acquired on-the-job on occupational assignment.

Having said this, it seems worth re-emphasizing that in our estimates years of schooling have by far the greatest influence on occupational assignment by skill level. Very large differences in work experience and in literacy score do not affect occupational assignment by as much as does a single year of schooling. In our view, some part of the effects of schooling are due to skills other than literacy skills which are produced by schooling. Nothing in our results allows us to say what skills these are and how much of the effect of schooling on occupational assignment is due to these skills.

Our second set of conclusions concerns the existence of returns to over-education and under-education. To our knowledge, the only other attempt to measure returns to over-education and under-education for Canada is Vahey's (2000) study, which uses data from a survey conducted in 1982. We used two different measures of required education, both based on the detailed occupation of persons in the IALS sample who are employed full-time. One of these measures shows our sample to be over-educated on average; the other shows it to be under-educated on average.

Our results using each of these measures of required schooling lead to the same conclusions. We find that a year of over-education earns a net positive return (up to 17 or more years of schooling) but that this return is substantially less than the return to a year of schooling when schooling matches the job requirements. Each year by which required schooling exceeds an individual's completed schooling (year of under-education) pays a return, but this return is less than the return

implied by having the same job and an additional year of schooling. These are the standard findings of the research literature on returns to over-education and under-education as summarized by Hartog (2000). They serve to emphasize the point that earnings depend crucially on the match between schooling and occupation, not on schooling alone.

As we have pointed out above, literacy skills are a significant determinant of occupational assignment. Our next set of conclusions concerns our attempt to assess the extent to which the returns to over-education and to under-education are determined by literacy skills. To do so, we included measures of literacy skills and of literacy skills and literacy use at work in our regression models for the returns to over-education and under-education.

When measures of literacy skills are included, the estimated coefficients of both over-education and under-education decrease in absolute value for men and the estimated coefficients of under-education decrease for women. Including both the measure of literacy skills and the measure of literacy use leads to further reductions in all of these coefficients, including those of over-education for women.

These results are consistent with the following model (discussed in Section 1.3 above). Workers with below average levels of literacy skills for their level of schooling are likely to be found in occupations whose schooling requirements are lower than these workers' level of education. They are likely to have above average literacy skills for these occupations and may be more likely to hold jobs in these occupations with a high level of literacy-related activities. Conversely, workers with above average levels of literacy skills for their level of schooling are likely to be found in occupations whose schooling requirements are higher than these workers' level of education. They are likely to have below average literacy skills for these occupations.

The higher earnings of the over-educated (relative to the earnings of those with just the required education) are in part a return to the literacy skills of the "over-educated", which are above average for the occupations in which they work. The lower earnings of the under-educated (relative to the earnings of those with just the required education) are a penalty to the literacy skills of the "under-educated", which are below average for the occupations in which they work. Their higher earnings (relative to those with the same schooling who work in occupations which require exactly this schooling) are in part a return to their higher literacy skills.

In our opinion, our results provide reasonably strong evidence for this model as concerns under-education for both women and men, since the coefficient of under-education in the regressions of Table 5 decreases when the measure of literacy skills is included as an explanatory variable. For over-education, the evidence for this model is strong for men (the absolute value of the coefficient decreases when the literacy skills measure is included) and weak for women (coefficient unchanged). For both women and men, the return to a year of over-education is substantially less than the return to a year of required schooling, even when literacy skills and literacy use are taken into account.

Literacy skills are not the only type of skill acquired through schooling and on-the-job training. As with literacy skills, there is likely to be variance among individuals with similar levels of schooling and work experience in their levels of these other skills. It is possible that if we had direct measures for these other skills, as we do for literacy skills, we could account completely for the returns to over-education and under-education.

A final, more speculative point is that our results seem to us to indicate that employers are capable of detecting differences in literacy skills within a level of schooling (or in other desired skills that are highly correlated with literacy skills). It would seem difficult to understand otherwise why persons with lower levels of literacy skills are assigned to occupations with below average skill requirements for their level of schooling and are paid less than average for the level of skill required in their occupation. This observation has implications for whether there are significant information failures in the labour market concerning employers' abilities to assess the skills of their employees or potential employees. Our results argue against the significance of such failures in the case of literacy skills.

References

Boothby, D. (1999) "Literacy Skills, the Knowledge Content of Occupations and Occupational Mismatch", Working Paper W-99-E3, Ottawa: Applied Research Branch, Human Resources Development Canada.

Borghans, L. and de Grip, A. (2000) "The Debate in Economics about Skill Utilization" in Borghans, L. and de Grip, A., *The Over-Educated Worker? The Economics of Skill Utilization*, Cheltenham: Edward Elgar, 3-23.

Boyd, M. (1990)" Sex Differences in Occupational Skill: Canada, 1961-1986", *Canadian Review of Sociology and Anthropology*, 27(3), 285-315.

Carey, S., Bridgwood, A. and Thomas, M. (2000) *Measuring Adult Literacy: The International Adult Literacy Survey in the European Context*, London: Office for National Statistics.

Charette, M. and Meng, R. (1998) "The determinants of literacy and numeracy, and the effect of literacy and numeracy on labour market outcomes", *Canadian Journal of Economics*, 31(3), 495-517.

DiNardo, J. and J. Pischke, (1997) "The Returns to Computer Use Revisited: Have Pencils Changed the Wage Structure Too?", *Quarterly Journal of Economics*, 112(1), 291-303.

Green, D. and Riddell, W. (2001) *Literacy, Numeracy and Labour Market Outcomes in Canada* (89-552 n8), Ottawa: Statistics Canada.

Green, F., McIntosh, S. and Vignola, A. (2000) "Overeducation: A tough nut to crack", *Centre Piece*, 5(1)

Hartog, J. (2000) "Over-education and earnings: where are we, where should we go?", *Economics of Education Review*, 19(2), 131-147.

Hunter, A. and Manley, M. (1986) "On the task content of work", *Canadian Review of Sociology and Anthropology*, 23(1), 47-71.

Kapsalis, C. (2000) *Literacy Profile of Ontario's Immigrants*, Toronto: Queen's Printer for Ontario.

Krahn, H. and Lowe, G. (1998) *Literacy Utilization in Canadian Workplaces* (89-552 n4), Ottawa: Statistics Canada.

Livingstone, D. (1999) *The Education-Jobs Gap: Underemployment or Economic Democracy*, Toronto: Garamond Press.

Manpower and Immigration (1971) *Canadian Classification and Dictionary of Occupations: Volume 1, Classification and Definitions*, Ottawa: Manpower and Immigration.

Manpower and Immigration (1973) *Canadian Classification and Dictionary of Occupations: Volume 2, Occupational Qualification Requirements*, Ottawa: Manpower and Immigration.

Murray, T., Kirsch, I. and Jenkins, L. (1998) *Adult Literacy in OECD Countries: Technical Report on the First International Adult Literacy Survey*, Washington: U.S. Government Printing Office.

Myles, J. (1988) "The expanding middle: some Canadian evidence on the deskilling debate", *Canadian Review of Sociology and Anthropology*, 25(3), 335-364.

Osberg, L. (2000) *Schooling, Literacy and Individual Earnings* (89-552 n7), Ottawa: Statistics Canada.

Pryor, F. and Schaffer, D. (1997) "Wages and the University Educated: A Paradox Resolved", *Monthly Labor Review*, 120(7), 3-14.

Pryor, F. and Schaffer, D. (1999) *Who's Not Working and Why: Employment, Cognitive Skills, Wages and the Changing U.S. Labor Market*, Cambridge: Cambridge University Press.

Sicherman, N. (1991) ""Overeducation" in the Labor Market", *Journal of Labor Economics*, 9(2), 101-122.

Statistics Canada (undated) *International Adult Literacy Survey: Canadian Data: Microdata User's Guide*, Ottawa: Statistics Canada.

Statistics Canada (1981) *Standard Occupational Classification: 1980* (12-565E), Ottawa: Statistics Canada.

Vahey, S. (2000) "The great Canadian training robbery: evidence on the returns to educational mismatch", *Economics of Education Review*, 19(2), 219-227.

Yamamoto, K. and Kirsch, I. (1998) "Proficiency Estimation", Chapter 11 in Murray et al (1998), 179-191.

Appendix 1

Sample Restrictions and Effects on Sample Size

Initial IALS sample: 2,423 men + 3,237 women = 5,660 persons

Restrictions applicable to all samples: no students or immigrants:

Men:	2,423 men – (315 students, non-immigrants + 183 immigrants, non-students + 20 immigrants, students = 1,905 men
Women:	3,237 women – (397 students, non-immigrants + 264 immigrants, non-students + 28 immigrants, students) = 2,548 women

Restrictions applicable to all samples in main body of text:

Required to report employment (at survey or during the previous year):

Men:	1,905 men – 600 not employed = 1,305 men
Women:	2,548 women – 1,243 not employed = 1,305 women

Required to report a valid 4-digit (unit group) 1980 Standard Occupational Code for major job:

Men:	1,305 men – 131 without valid SOC code = 1,174 men
Women:	1,305 women – 34 without valid SOC code = 1,271 women

1980 SOC code required to match to a CCDO unit group (valid GED and SVP):

Men:	1,174 men – 64 unable to match = 1,110 men
Women:	1,271 women – 15 unable to match = 1,256 women

Required to work full-time at main job:

Men:	1,110 men – 61 part-time = 1,049 men
Women:	1,256 women – 341 part-time = 915 women

The last two lines describe the sample for Table 1. In this sample, 3 men and 1 woman were missing the years of completed schooling variable and an additional 11 men and 14 women were missing the highest level of studies variable.

It is also the initial sample for the regressions in Tables 2 and 3. In these regressions, missing values for the regressors led to 124 additional missing observations for men (leaving a sample size of 925 men) and 79 additional missing observations for women (leaving a sample size of 836 observations).

This is also the initial sample for the regressions in Tables 4 and 5. All of the suppressions due to missing values in Tables 2 and 3 apply to Tables 4 and 5. In addition, 197 men and 198 women who were included in the samples of Tables 2 and 3 did not report their earnings. This further reduced the sample sizes to 852 men and 717 women. Finally, the literacy use variable could not be constructed because of missing responses for 78 men and for 33 women, resulting in final sample sizes in Tables 4 and 5 of 774 men and 684 women.

For the estimates reported in Appendix 3, we begin with the population of 1,905 men + 2,548 women = 4,453 persons who are neither students nor immigrants. In Appendix Table A.1 we report correlations whose sample sizes vary due to missing responses. Among the 4,453 persons of the initial sample, there are no missing responses for the literacy score, there are 788 persons who did not take and complete the second part of the test, and there are 27 persons for whom the years of schooling completed are missing Of these 27, 14 also did not take and complete the second part of the literacy test.[22] The smallest group for the correlations is thus for the correlation of years of schooling with the percentage of right answers (4,453 – 788 – (27 – 14)) = 3,652. This is also the sample for the regressions reported in Appendix Tables A.2 and A.3.

Appendix 2

Issues Concerning the Functional Form of the Earnings Equations

Non-Linearity in the Return to Schooling

As shown in Tables 1, 2 and 3, as years of completed schooling increase, over-education increases (under-education decreases). Suppose that the return to schooling decreases as years of schooling increase (declining marginal returns). Suppose further that a linear specification of the return to schooling is imposed, and that measures of over-education and under-education are included in the specification. The estimated coefficients of over-education and under-education will capture the non-linearity of the return to schooling, and thus will tend to show returns to over-education and to under-education which may be spurious.

For this reason, previous researchers into the returns to over-education and under-education have tested for non-linearity in the returns to schooling. Hartog (2000, 135) cites several studies whose authors found either increasing returns to schooling or no evidence of non-linearity of the return to schooling.

This is not the case with the sample we used from the IALS. As shown in Tables 4 and 5, the linear term is positive and the quadratic term is negative when we use a quadratic specification of the returns to schooling. We therefore account explicitly for the non-linearity of the returns to schooling by using a quadratic specification for schooling throughout.

We nonetheless find statistically significant effects of over-education and under-education. We experimented with functional forms for the earnings equations with linear, quadratic and cubic terms in years of schooling. When both cubic terms in schooling and linear measures of over-education and under-education were included in the specification, over-education and under-education continued to have statistically significant effects, while the cubic term (or occasionally the linear term) in years of schooling was not statistically significant.

We conclude that there are both non-linearities in the return to schooling and returns to over-education and under-education in our sample. Of course we cannot entirely rule out the possibility that the measured returns to over-education and under-education are artifacts due to the functional form of the return to schooling.

Bias in the Measure of Over-Education and Under-Education

We now consider the effects of biased measures of over-education and under-education. Suppose, for example, that we consistently under-estimate required schooling by one year. Suppose further that the correct model for earnings includes measures of over-education O and under-education U with their respective coefficients o and u. Using O and U to represent the true levels of over-education and under-education, we can rewrite the model of equation 3 as:

$$3'. \ln Y = c + sS + rS^2 + [o(O-1) + od_O] + [u(U+1)-ud_U] + X'B + e$$

where d_O and d_U are indicator variables for over-education and under-education respectively.

Our under-estimation of required schooling leads us to define over-education as $O' = \{S-(R-1), S-(R-1) > 0; 0 \text{ otherwise}\}$ and under-education as $U' = \{(R-1)-S, (R-1)-S > 0; 0 \text{ otherwise}\}$. No problem arises if $o = -u$.

If this relation does not hold, there are two problems. First, everyone who is correctly identified as over-educated ($O' > 1$) will have measured over-education O' greater than actual over-education O by one year, while everyone who is correctly identified as under-educated ($U' > 0$) will have their under-education under-estimated by a year. This will change the regression constant (by a weighted average of the estimates of u and o in the correct specification).

The second and more serious problem is that all individuals for whom $1 > O' > 0$ will be identified as over-educated, when they are in fact under-educated. This will bias estimates of o, since the estimated effect of the difference between schooling and required schooling for these individuals is forced to the estimate of o when it should be u.

Considerations like these led us to the use of two measures of over-education and under-education. One of these, based on GED as a measure of required schooling, shows over-education on average. The other, based on GED+SVP as a measure of required training time shows under-education on average. Both give qualitatively similar results in Table 5. These results agree with the usual results of estimates of over-education/under-education models. Since we have no grounds for knowing which is a better measure of required schooling, we do not know which are the better point estimates of the effects of over-education and under-education.

The problem as posed is that we do not know where over-education ends and under-education begins, but our functional form constrains us as to where we pass from the one to the other. We carried out alternative estimates for a specification which segmented the difference R-S and allowed different coefficients for each segment. If the model of equation 3 is correct, the return to R-S in each segment of over-education would be –o, the return to R-S in each segment of under-education would be u and the return in the segment in which the change-over from over-education to under-education occurred would lie somewhere between the two.

In these estimates large values of R-S had large, statistically significant effects, while small values of R-S tended not to have statistically significant coefficients. We were not able to identify a segment where the changeover from over-education to under-education occurred, as there were usually several adjacent segments with small values of R-S where the coefficient of R-S was not statistically significant. (For example, for men, for both the GED and GED+SVP measures of R, the coefficients were not statistically significant for the segments [-2,-1), [-1,0), [0,1)).

We would interpret these results as indicating that R-S is measured with considerable error, leading to lack of precision in estimates of its effects for small values of R-S. It may be that one or both of our measures of R-S are biased, but the results of these regressions do not allow us to identify the size or even the direction of the bias.

Another possible difficulty with the specification of equation 3 is its assumption that the effect of differences between schooling required by the individual's job and schooling completed by the individual is piecewise linear (with the kink at 0). Other types of non-linearity are possible. We tested this by estimating quadratic specifications in R-S (with returns to S also estimated in quadratic form). The linear and quadratic terms in S were always statistically significant in this specification.

The only case in which the quadratic term in (R-S) was statistically significant was for women, when the required measure of schooling was GED+SVP. In this case, the partial derivative of earnings with respect to R-S was positive for R-S<12, negative thereafter.

We saw no strong reason in any of these results to prefer either of these functional forms to the more conventional forms we reported in Tables 4 and 5.

Appendix 3

The Relation between Education, Literacy and Literacy Scores

In Tables 3 and 5, we included two measures of literacy skills, the literacy score (average of the first plausible values for prose, document and quantitative literacy) and the percentage of correct answers on the literacy exams. We noted that the literacy score is conditioned on education, that there are differences between the effects of the two literacy measures on occupation, and that these differences are generally in the direction of smaller effects for the percentage correct measure than for literacy score. Since the qualitative results with the two measures are similar, we reported the more widely used literacy score.

This appendix further explores the relation between literacy and education and the effects of conditioning the literacy score on "background variables". A review of the IALS published by the British Office of National Statistics for the European Commission states that "...there is a concern that the relationships observed between background characteristics and the skill levels are exaggerated by the IRT procedures where background characteristics are used to condition the estimates. One questions the validity of using multivariate techniques for example on the proficiency estimates as one is never sure how much of the perceived relationship is due to the factors in the analysis having already been taken into account in the conditioning and estimation procedures. Of particular concern is the potential for identifying spurious relationships." (Carey et al, 2000, 245) In Chapter 7 of the same report, Heady argues for using more easily understood percentages of correct responses.

Construction of "Plausible Values" in the IALS

We present below an empirical examination of the effects of conditioning of literacy scores on the measured influence of education on literacy. Before giving our results, we need to discuss how the literacy scores were constructed. What follows is our interpretation of Yamamoto and Kirsch (1998).

The outcome of the IALS assessment of literacy was five plausible values for each of three types of literacy—prose literacy, document literacy and quantitative literacy. (Each respondent was also assigned to one of four levels for each type of literacy, but this will not concern us here). The measure of literacy we called the literacy score is the average of the first plausible value for each of the three types of literacy. (There does not seem to be any implication that the first plausible value is more plausible than any of the others).

Yamamoto and Kirsch (1998, 180) state very clearly why the method of plausible values is adopted. "It cannot be emphasized too strongly that plausible values are not test scores for individuals in the usual sense. In other words, they are unlike the more familiar ability estimates of educational measurement which are, in some sense, optimal for each respondent (e.g., maximum likelihood estimates and Bayes estimates). Point estimates that are optimal for individual respondents can produce decidedly nonoptimal (inconsistent) estimates of population characteristics (Little and Rubin, 1983). Plausible values are constructed explicitly to provide consistent estimates of population effects, even though they are not generally unbiased estimates of the proficiencies of the individuals with whom they are associated (Mislevy, Beaton, Kaplan and Sheehan, 1992)".[23] Since our study (and others) used these scores as individual-level measures, it seems worth looking somewhat further into how they were produced.

In the methodology described, each type of literacy is treated as an unobserved latent variable. The "scale proficiency values" (latent literacy levels) are not observed for the sampled respondents, so it is not possible to construct a statistic t (for example, a sub-population mean) of the scale proficiencies. Instead, the expectation of t given the data actually observed, t*(x,y), is estimated. Here, the x variables are item responses and the y variables are background variables. According to Yamamoto and Kirsch (1998, 180), "It is possible to approximate t* using random draws from the conditional distribution of the scale proficiencies given the item responses x_j, background variables y_j and model parameter for sampled respondent *j*. These values are referred to as imputations in the sampling literature and as plausible values in the IALS".[24] The process is repeated to allow estimation of the uncertainty due to the fact that scale proficiencies are not observed.

Plausible values for each respondent *j* are drawn from the distribution of the latent scale proficiency values, conditional on x_j, y_j, a matrix of regression coefficients G, and S, a common variance matrix for residuals. This conditional distribution is computed from the "product over the scales of the independent likelihoods induced by responses to items [the x_j] within each scale" and "the multivariate joint density of proficiencies of the scales, conditional on the observed value y_j of background responses and parameters G and S. Item parameter estimates are fixed and regarded as population values in the computation...". (Yamamoto and Kirsch, 1998, 181)

The multivariate joint distribution of the three proficiencies was assumed multivariate normal "with a common variance, S, and mean given by a linear model with slope parameters, G, based on the first approximately principal components of several hundred selected main effects and two-way interactions of the complete vector of background variables. The background variables included sex, ethnicity, language of interview, respondent education, parental education, occupation and reading practices, among others. Based on the principal component method, components representing 99 percent of the variance present in the data were selected. " These principal components are the conditioning variables used in the analysis." (Yamamoto and Kirsch, 1998, 181). The authors then describe the algorithms used to estimate G and S and how the plausible values can be used to derive estimates of statistics of the latent scale proficiencies.

Estimating the Effects of Conditioning

At this point, we would make the following comments. First, all background variables (anything which is not a test response) are used in constructing the literacy measures. It is thus open to question how much of the observed relation between any "background variable" for an individual respondent and the plausible values for the individual respondent are due to this conditioning. Second, it would be very difficult to trace through the influence of any "background variable" on individual plausible values.

Consequently, we do not attempt to trace through the influence of education on the literacy score (average of the three first plausible values) via the conditioning of the plausible values on background variables. Instead, we compare the effects of education on the literacy score and on an

alternative measure of literacy skills (the percentage of right answers) in simple regression models. We also show that educational attainment has a significant effect on the literacy score, even when the percentage of right answers is taken into account.

We measure the proportion of correct responses only for persons who were coded as having completed the literacy questionnaire. This excludes persons who completed the background questionnaire but refused the literacy assessment, persons who did not answer at least two items correctly in the screening questionnaire (core booklet) and persons who refused to complete the assessment at some time during the assessment. Thus the population for whom we computed the percentage correct excludes all of the population for whom imputation procedures were required to compute the plausible values and all of the population whose literacy skills were so low that they were unable to pass the level of the screening questionnaire.

We computed the proportion of correct responses as the number of correct answers/(number of correct answers + number of incorrect answers + number of questions refused/not done). Note that this has the effect of treating all refused and omitted items as wrong answers. Both the screening questionnaire and the questions from the test booklets were included in this calculation. Only the scoring by the first scorer was used (the IALS used two scorers for 20% of respondents).

The mean proportion of correct responses (for persons in our sample for whom we could carry out this computation) was .683. Yamamoto and Kirsch (1998, 83) report mean percentages correct for the prose, document and quantitative literacy scales respectively of 70%, 74% and 67% for Canada/English and of 64%, 69% and 60% for Canada/French.

The mean value of the literacy score measure (average of first plausible values) for the same sample was 282. Yamamoto and Kirsch report mean proficiency values for the prose, document and quantitative literacy scales respectively of 284, 284 and 286 for Canada/English and of 264, 265 and 266 for Canada/French. The mean value of the literacy score measure in our non-student, non-immigrant sample is 272. The lower value of the literacy score for this broader sample presumably reflects the low plausible values assigned to persons who did not advance beyond the screening questions and to many of those who refused the test.

Appendix Table A.1 gives the correlations between the literacy score, the proportion of correct responses and years of schooling. There are two points to note. First, the correlation between the literacy score and the proportion correct (.88) is quite high. Second, the correlation between the literacy score and schooling is much higher (.70) than the correlation between the proportion correct and schooling (.58).

Table A.1	Correlations among Literacy Score, Proportion Correct and Years of Schooling		
	Literacy Score	Proportion Correct	Years of Schooling
Literacy Score	—	0.88 n = 3,665	0.70 n = 4,426
Proportion Correct	—	—	0.58 n = 3,652

Several authors have attempted to estimate the influence of schooling on literacy skills using the literacy score or a similar variable as the measure of literacy skills. Appendix Table A.2. presents the results of two sets of regression estimates with literacy skills measures as the dependent variable. In the first the literacy score is the dependent variable, in the second the proportion of correct answers is the dependent variable. Both dependent variables have been normalized to have mean 0 and standard deviation 1. This allows direct comparison of the estimated coefficients across the two regressions.

These models should not be interpreted as attempts to measure the structural relation between schooling and literacy. In a cross-section such as the IALS, literacy skills and attained schooling are determined simultaneously. Some or all of the correlation between schooling and literacy skills may occur because literacy skills act as a screen for further schooling, not because further schooling produces literacy skills.[25] In our opinion, the structural effects of schooling on literacy can best be measured by repeated measurements of individuals' literacy skills at different points in their schooling.

Table A.2   **Regression Estimates for Literacy Score and Proportion Correct**

	Literacy Score		Proportion Correct	
Constant	*-.0386*	*(.0019)*	*-.0365*	*(.0020)*
Years of School	*.0028*	*(.0001)*	*.0025*	*(.0001)*
Female	.0005	(.0004)	**.0011**	**(.0004)**
Age	*.0004*	*(.0001)*	*.0006*	*(.0001)*
Age^2	*-6x10^{-6}*	*(8x10^{-7})*	*-8x10^{-6}*	*(8x10^{-7})*
Tested in French	*-.0032*	*(.0005)*	*-.0038*	*(.0005)*
Not Tested in Native Language	*-.0039*	*(.0008)*	*-.0033*	*(.0008)*
r^2 corrected for d.o.f.	.45		.41	

n = 3,652

Estimated standard errors in parentheses.

Dependent variables standardized to mean = 0 and S.D. = 1

Bold-face and italic: Significant at 1%. Bold-face only: Significant at 5%. Italic only: Significant at 10%.

The coefficients of the regressors in the two models of Table A.2 are very similar, with one exception. The effect of gender on the literacy score is not statistically significant, while the effect of gender on the proportion of correct responses is statistically significant. This effect is not large, however, amounting to less than half the effect of a completed year of school.

Taking the test in French or taking the test in a language other than one's native language have negative, statistically significant effects on both the literacy score and the percentage of right answers. Almost all of the persons who took the test in a language other than their native language were persons of French native language who took the test in English.

While there is some difference between the coefficients of the terms in years of age, the differences between the partial derivatives of the dependent variables with respect to age are small in the age range 20 years to 60 years (the largest differences is 8.3 x 10^{-5} at age 60 years). The partial derivative of the literacy score with respect to age is positive through age 33.5 years, negative thereafter. The partial derivative of the proportion of correct responses with respect to age is positive through age 34.1 years, negative thereafter).

All of the regressors in these two models are among the background variables used in establishing the literacy score. There is no evidence in these regressions that any of them influence literacy scores more than they influence the percentage of correct responses. In particular, the conditioning of literacy scores does not seem to have exaggerated the relation between completed schooling and literacy skills.

The correlations in Table A.1 show that the literacy score variable is closely related to the percentage of correct responses. We now examine the relation between the percentage of correct responses and the literacy score in a multivariate context. Appendix Table A.3 gives the results of regressing the literacy score on a cubic in the proportion of correct responses, the regressors included in Table A.2 and a series of dummy variables that identify which of the seven versions of the test the person took.

Regressors which affect the literacy score only because they affect the percentage of correct responses should not have statistically significant effects in this regression. We entered the percentage of correct responses as a cubic because all three terms of the cubic expression are statistically significant, and we seek to account as completely as possible for the effect of correct responses on the literacy score. Statistically significant effects of "background variables" should be an approximation of their effect via conditioning on the literacy scores of persons in this sample.[26]

Table A.3   Regression Estimates for Determinants of Literacy Score

	Literacy Score	
Constant	*-.0103*	*(.0011)*
Years of School	*.0007*	*(4x10⁻⁵)*
Female	- .0003	(.0002)
Age	-1×10^{-5}	(4×10^{-5})
Age^2	-9×10^{-8}	(4×10^{-7})
Tested in French	.0002	(.0003)
Not Tested in Native Language	*-.0013*	*(.0004)*
Proportion Correct	*.8289*	*(.0133)*
Proportion Correct^2	*17.30*	*(.6869)*
Proportion Correct^3	*277.1*	*(20.3500)*
Test 1	*-.0025*	*(.0004)*
Test 2	*.0019*	*(.0004)*
Test 3	*-.0023*	*(.0004)*
Test 4	*-.0024*	*(.0004)*
Test 5	*-.0014*	*(.0004)*
Test 6	*-.0022*	*(.0004)*
Test 7	—	—
r^2 corrected for d.o.f.	0.84	
n = 3,652		

Estimated Standard Errors in Parentheses

Dependent variable standardized to mean = 0 and S.D. = 1

Bold-face and italic: Significant at 1%. Bold-face only: Significant at 5%. Italic only: Significant at 10%

Gender, the terms in age and the language of the test do not have statistically significant effects on the literacy score in the regression of Table A.3. The effect of years of schooling completed is statistically significant. Each additional year of school is predicted to increase the literacy score by slightly over two points. (Computed as the coefficient x 3154, the standard deviation of the literacy score in the sample before standardizing). The effect of taking the test in other than the native language is also statistically significant, reducing the literacy score by four points on average.[27]

The version of the test taken has statistically significant effects on the literacy score. This is undoubtedly due to differences in the difficulty of the versions which are not captured by the proportion of correct responses. The proportion of correct responses measure would be improved if it were adjusted for the test version.

The r^2 for the model (.84) shows that the variables included account for most of the variance in the literacy score.

Some Conclusions

Two of the background variables included in the model of Table A.3 have statistically significant effects on the literacy score, once the percentage of correct responses to the test items and the effect of the test version are taken into account. One of these is the widely used variable years of completed schooling. One must conclude that for important variables on the IALS file, there is indeed a "potential for identifying spurious relationships" (Carey et al, 2000, 245) or at least a potential for exaggerating the size of the effects in the relationship.

The effects of the background variables in Table A.3 are small. Moreover, the results in Table A.2 indicate that various regressors have very similar effects on the literacy score and on the proportion of correct responses. There may be very little danger in using the plausible values as measures of literacy skill when estimating structural relationships, if one limits the sample to persons who completed the literacy test. Researchers are likely to continue to be uneasy with the plausible values, however, especially in the face of differences in results between plausible values and proportion correct like those found in Tables 4 and 5.

One can wonder why the plausible values should be the only continous-valued measures of literacy skills which appear on public microdata files. The point of having a microdata file is in large part to allow research at the individual level. Yamamoto and Kirsch (1998, 180) explicitly state, however, that the plausible values "are not generally unbiased estimates of the proficiencies of the individuals with whom they are associated". In our opinion, the public microdata files should contain unbiased estimates of the proficiencies of individuals.

The proportion of correct responses is a crude method of constructing such a measure. This measure could be produced in a version in which the effect of the test version was corrected. A better approach might be to use the item difficulty ratings for each literacy scale to construct individual proficiencies for each literacy type. The item difficulty ratings are published in the Appendix materials for Chapter 11 (Yamamoto and Kirsch, 1998) in Murray et al (1998).

A measure of this type can only be constructed for persons who completed the literacy test. In our opinion, the principal danger in the use of the plausible values is not the small effects of the background variables that we detected. It is the use of the methodology to produce "plausible values" for individuals who did not take the literacy test.

Yamamoto and Kirsch (1998, 190) justify the reasons for the treatment of non-response in terms of the need for unbiased proficiency estimates for populations and sub-populations. One can have doubts about the adequacy of these methods. For example, these methods are presumably applicable only to persons who have completed the background questionnaire. Why should total refusal (including for the background items) be any less likely to produce biased results than refusal for all or part of the literacy tests?

Whatever one thinks of the methods used with regard to the need to produce unbiased estimates for populations, there is no reason to think that the plausible values produced by these methods are accurate measures for individuals. They are treated as such by researchers using the public microdata files, however. To our knowledge, there is no way to identify in the public microdata files those persons for whom the plausible values reflect both test scores and background variables and those for whom the plausible values are based entirely on background variables.

In our opinion the inclusion in individual-level estimations of individuals whose plausible values are based entirely on background variables is the likeliest source of "identification of spurious relationships". We think that at a minimum, public use data files should allow researchers to distinguish individual plausible values based only on "background characteristics" from those which reflect test results. Preferably, the public use data should contain a valid measure of individual proficiencies, available only for those who completed the literacy test.

Endnotes

1. Note that R, O and U decompose S. All coefficients equal therefore amounts to equal returns to all years of schooling, regardless of the educational requirements of the job. Hartog also notes that various authors have established that this result is not due to non-linearity of the return to schooling. These points are discussed further in Appendix 2.

2. According to Vahey (2000,220) the *National Survey of Class Structure and Labour Process* was a cross-sectional survey with about 3,000 respondents interviewed in 1982 by Canada Facts. Their responses were decoded and transferred to tape at the Department of Sociology and Anthropology at Carleton University. Since the survey collected the respondent's highest level of completed schooling and the level of schooling required for the job, Vahey's measures of over-education and under-education are yes/no indicators rather than the continuous measure I described above. Various other authors also use dichotomous indicators for over-education and under-education.

3. The LSUDA did not collect respondent earnings. Charette and Meng (1998) use previous year income for persons who worked in the previous year.

4. The definition of "mismatch" used is working outside the "skilled information" category of occupations.

5. Appendix 1 provides a complete list of sample restrictions and their effect on sample size. The sample restrictions (in particular, the restriction to full-time workers) may not be random with respect to literacy skills.

6. There are minor differences between the CCDO and the 1980 SOC at the unit group level. We reassigned CCDO unit groups to 1980 SOC unit groups based on the concordance between the CCDO and the OCM described in the 1980 SOC manual (Statistics Canada, 1981, 14-18).

7. These levels of GED and SVP in years for detailed occupations were developed by Wayne Roth, Human Resources Development Canada from the GED and SVP levels in the CCDO, using the method described in the text.

8. If one asks a hundred employees with the same job title and the same employer the educational requirements for their job, there is likely to be variance in the responses, even though the requirements for these jobs are exactly the same. There is also no guarantee that the mean response will be equal to the actual requirement.

9. They attribute this concept to J.B. Knight. The idea of an education-wage profile (or a skill-wage profile) in an occupation would seem to introduce some uncertainty as to what constitutes the "schooling requirements" or the "skill requirements" of an occupation. One possibility is that these requirements represent a minimal level of skills needed to function in the occupation. Borghans and de Grip discuss an optimal skill match in the occupation based on the relation between the education-wage profile determined in the labour market and the education-productivity profile unique to the occupation. Over-educated workers in an occupation would have higher productivity and be paid more than less-educated workers.

10. See Green and Riddell (2001) for a more formal presentation of this argument in the context of the coefficients of schooling and literacy in an earnings equation.

11. Quadratic terms in schooling and work experience did not have statistically significant effects in preliminary regressions and are omitted from the models reported in Table 2.

12. On the other hand, the coefficient to work experience, thus estimated, is greater in the earnings equation for women than for men. (Table 4) If the measure of experience is biased upward for women, the results of this bias are more than offset by greater earnings returns to experience for women. Obviously it would be preferable to have a direct measure of work experience and to have a measure of job tenure. Neither is available in the IALS.

13. In the initial stages of our work we used provincial variables, both here and in the wage estimations of the next section. Differences among the estimated coefficients among provinces within a region were small, and the sample in some provinces was small. For these reasons, we preferred to report coefficients for regional variables.

14. The IALS data file provides five scores, called the first plausible value, second plausible value, etc. for each of the three types of literacy measured (prose, document and quantitative). Each score is scaled from 1 to 500 (with proficiency increasing as the score increases). Green and Riddell (2001) provide what seem to us convincing

arguments for using the sum or the average of the scores for the three types of literacy. They use the average (presumably of the first plausible values). Osberg (2001) uses the sum of the first plausible values for the three scores. In our sample for the regressions in Appendix 3 (n = 3,652), the mean of the average literacy score is 282, the mean of the proportion of correct answers is .68 and the Pearson correlation of these two measures is .88.

15. The results for the other variables included in Table 2 are very similar to those shown in Table 2 and are omitted in Table 3.

16. We evaluate the marginal returns to schooling and to work experience by taking the partial derivatives of our estimated model with respect to these variables. Thus the marginal return to a year of schooling is the coefficient of the linear term in schooling plus twice the coefficient of the quadratic term in schooling times the years of schooling, and similarly for work experience.

17. The coefficients of the other variables shown in Table 4 changed little in the regressions underlying Table 5 and are not shown.

18. Sicherman (1991) also discusses the functional form of equation 2 and the identities linking the coefficients of the functional forms of Equations 1 and 2. Appendix 2 provides additional discussion of issues of functional form.

19. The reading activities are the responses from "How often (do/did) you read or use information from each of the following as part of your main job?": 1. Letters or memos, 2. Reports, articles, magazines or journals, 3. Manuals or reference books, including catalogues, 4. Diagrams or schematics, 5. Bills or invoices 6. Directions or instructions for medicines, recipes or other products. We did not use the seventh reading item "material written in a language other than English (French)". The writing activities are the responses from "How often (do/did) you write or fill out each of the following as part of your main job?": 1. Letters or memos, 2. Forms or things such as bills, invoices or budgets, 3. Reports or articles, 4. Estimates or technical specifications. We scored the responses to each item to increase from 0 to 4 as the frequency of the activity increased from "rarely/never" to "every day".

20. We attempted to test whether the effects of literacy skills and literacy use at work on earnings occur only when the individual who has these skills uses these skills. When an interaction variable is included in the earnings regression, its coefficient is positive and it has a high level of statistical significance. When the literacy skills measure, the literacy use measure and their interaction are all included, the results are difficult to interpret. We therefore chose to report the functional form which includes the literacy skill and literacy use measures, but not their interaction.

21. More exactly, since the return to schooling declines as schooling increases, this is only true up to the value of years of schooling at which the marginal return to schooling is equal to the penalty to over-education. For men, this threshold is attained at 18 years of school in the model of column 2 and 16 years of school in the model of column 3. The corresponding levels for women are 18 years of school and 17 years of school.

22. The first part of the test consists of six screening questions. Persons who failed to answer at least two of these six questions correctly were not administered the second part of the test, which consisted of a larger number of questions of varying levels of difficulty.

23. The citations in the quote are given in the References to Murray et al (1998) as Little, R.J.A. and Rubin, D.B., 1983, "On jointly estimating parameters and missing data", *American Statistician*, 37, 218-220 and Mislevy, R.J., Beaton, A., Kaplan, B.A. and Sheehan, K., 1992, "Estimating population characteristics from sparse matrix samples of item responses", *Journal of Educational Measurement*, 29(2), 133-161.

24. Note that the statistic t of the scale proficiencies referred to in the quote is any statistic of the scale proficiencies and not in particular a (Student) t-statistic.

25. An example of this effect would occur if individuals' levels of the types of literacy skills tested in the IALS were completely determined by the end of high school. High school graduates with high levels of literacy skills would be more likely to continue on to post-secondary studies. This would lead to correlation between post-secondary schooling and literacy skills, but would not imply that post-secondary education leads to the acquisition of further literacy skills.

26. We emphasize that this is a sample of persons who completed the literacy test. The effects of the background variables on the literacy scores of persons who did not complete the test are likely to be much larger.

27. I am informed that almost all French native language individuals who took the literacy test in English came from the Franco-Ontarian special sample. Consequently, the effect here may be due to some other characteristic of this sample, and not to the fact of taking the test in other than the native language.

International Adult Literacy Survey

Monograph Series

The International Adult Literacy Survey (IALS) was a seven-country initiative conducted in the fall of 1994. Its goal was to create comparable literacy profiles across national, linguistic and cultural boundaries. Successive waves of the survey now encompass close to 30 countries around the world.

The Monograph Series features detailed studies from the IALS database by literacy scholars and experts in Canada and the United States. The research is primarily funded by Human Resources Development Canada. Monographs focus on current policy issues and cover topics such as adult training, literacy skill match and mismatch in the workplace, seniors' literacy skills and health, literacy and economic security, and many others.