Understanding Educational Impacts: The Role of Literacy and Numeracy Skills*

David A. Green
Vancouver School of Economics
University of British Columbia

W. Craig Riddell
Vancouver School of Economics
University of British Columbia

October 2015

Abstract

Much research concludes that formal education exerts a powerful causal influence on individual earnings as well as various non-pecuniary outcomes. However, little is known about the mechanisms that underlie these impacts. This paper investigates the extent to which the estimated impacts of formal schooling reflect the impact of education on the production of literacy and numeracy skills and the influence of these skills on earnings. To do so we use a rich data set containing measures of literacy, numeracy and problem-solving skills for a representative sample of the adult population.

Our investigation yields several noteworthy findings. We provide strong evidence that education has a substantial causal effect on cognitive skills, and that formal schooling is the dominant determinant of basic literacy, numeracy and problem-solving skills. In addition, our estimates imply that 35 to 40 % of the return to schooling represents the combined effect of education on the production of literacy and numeracy skills and the value placed on these skills in the labour market. This is a substantial component of the return to education, but its size suggests that schooling also affects earnings via other mechanisms. We also find that the direct effect of cognitive skills on earnings is substantial. A 25-point increase in literacy and numeracy skills (half of a standard deviation) is associated with an increase in earnings equivalent to an additional year of schooling.

Keywords: literacy, numeracy, problem-solving, cognitive skills, education, earnings, compulsory schooling laws

JEL codes: I20, J24, J31

1. Introduction

Education has numerous – and profound – consequences for individuals and society. As many studies have documented, education is one of the best predictors of "who gets ahead." Better-educated workers experience higher lifetime earnings, less unemployment and work longer. Higher education is also associated with longer life expectancy, improved health, reduced participation in crime, and greater civic participation. These correlations have been known for along time. A substantial body of recent research concludes that these relationships reflect, at least in part, causal influences of schooling on individual outcomes. The relationship between education and earnings has been extensively investigated. As the surveys by Card (1999, 2001) indicate, there is now strong evidence that schooling exerts a substantial causal impact on individual earnings. Similarly, recent research finds evidence of causal linkages between education and numerous non-pecuniary outcomes (Grossman, 2006; Oreopoulos and Salvanes, 2011). These include reduced participation in crime (Lochner and Moretti, 2004), greater civic participation (Dee, 2004; Milligan, Moretti and Oreopoulos, 2004) and greater life satisfaction (Oreopoulos, 2007).[1] There is also growing evidence of inter-generational impacts, implying that benefits also accrue to the children of those receiving additional schooling (e.g. Plug, 2004; Black, Devereux and Salvanes, 2005; Oreopoulos, Page and Stevens, 2006).

Although these and other consequences of additional schooling are increasingly becoming understood, much less is known about the mechanisms through which schooling exerts such powerful effects. Does schooling enhance individuals' cognitive skills such as literacy, numeracy and problem-solving skills, enabling them to acquire and process information as well as perform more complex tasks in the workplace, thus increasing their value to employers? As Sen (1999) emphasizes, individuals without basic literacy and numeracy skills cannot assume a full and equal role in social and political discourse. If schooling plays an important role in the production of basic literacy and numeracy skills, low earnings and low levels of civic participation could be a consequence of limited education. Or do the consequences of education arise because additional schooling enhances non-cognitive skills such as reliability, time management, and the ability to get along with others and to work effectively in teams? Another possibility is that additional education – to the extent that it is associated with higher earnings -- relaxes budget constraints. Improvements in health and well-being could be associated with higher income, to the extent that income is not taken into account in studies of the relationship between

---

[1] Evidence of causal impacts of education on health and longevity is mixed (Lleras-Muney, 2005; Oreopoulos, 2007; Clark and Royer, 2013; Grossman, 2015).

schooling and various non-pecuniary outcomes. Finally, education may alter individual preferences – making people more "forward looking" and thus more willing to make investments – such as adopting healthy behaviors – that pay off in the future.

The purpose of this paper is to investigate the extent to which the impacts of education arise from the effects of schooling on basic literacy and numeracy skills and the consequences of these skills for labour market outcomes. Specifically, we provide estimates of the fraction of the return to schooling that can be attributed to the combined effect of education on the production of literacy and numeracy skills and the value placed by these skills in the labour market.

Examining the role played by cognitive skills production seems a natural starting point for beginning to understand why schooling has powerful effects on many dimensions of individuals' lives. The concept of human capital has traditionally emphasized the acquisition of such skills as a key part of the educational process. Although education has multi-dimensional objectives, school systems around the world are judged, at least in part, by their ability to impart basic literacy, numeracy and analytical skills to students. Further, the importance of literacy, numeracy and problem-solving skills is often emphasized by employers.

Key to our investigation is a rich source of data on the literacy, numeracy and problem-solving skills of a representative sample of the adult population. These data also contain rich information on labour market outcomes, educational attainment and family background. Using these data we study the causal impact of education on literacy, numeracy and problem-solving test scores. We then estimate the impact of schooling on earnings using methods similar to those used in previous studies. Introducing measures of literacy and numeracy skills into these equations allows us to assess the extent to which estimated returns to schooling reflect the cognitive skill production mechanism.

Our analysis of the effect of schooling on skill production and the impacts of education and skills on earnings must deal with the potential endogeneity of educational choices. To do so we take advantage of the rich data and control for variables such as family background and ability proxies that are often unobserved. We also adopt an instrumental variables strategy using both Canadian compulsory schooling laws and information on the province in which the respondent attended secondary school to form instruments for education.

Our investigation yields several noteworthy findings. First, we provide strong evidence that education has a substantial causal effect on basic literacy, numeracy and problem-solving skills, and that formal schooling is the dominant determinant of these cognitive skills. Our instrumental variables

estimates indicate that each additional year of schooling raises average skills by about 7% or about one-tenth of a standard deviation of the skill score distribution. In addition, age (or experience) has little impact on literacy and numeracy skills, suggesting that the positive relationship between experience and earnings arises for other reasons. Furthermore, parental characteristics have only modest effects on cognitive skills, once we control for the individual's education. The influence of parental characteristics on skills arises indirectly through their powerful influence on the child's education.

Our estimates imply that about 35% to 40% of the return to schooling represents the combined effect of education on the production of literacy and numeracy skills and the value placed on these skills in the labour market. This is a substantial component of the return to education, but its size suggests that schooling also affects earnings via other mechanisms. We also find that the direct effect of cognitive skills on earnings is substantial. A 25-point increase in literacy and numeracy skills (half of a standard deviation) is associated with an increase in earnings equivalent to an additional year of schooling.

The paper is organized as follows. The next section outlines a simple conceptual framework that underlies our analysis. Section 3 describes our data and its suitability for studying the cognitive skill production mechanism, while section 4 presents estimates of the causal impact of formal schooling on literacy, numeracy and problem-solving skills. The fifth section examines the extent to which typical estimates of the returns to schooling reflect the combination of the effect of schooling on cognitive skills and the way these skills are rewarded in the labour market. The final section concludes.

2. Conceptual Framework

This section sets out a simple hedonic framework for considering earnings generation and its relationship to cognitive skills. Our approach builds on Welch (1969) and Rosen (1974). We distinguish between skills (personal characteristics that aid in productivity in specific tasks and that can be acquired by the individual) and abilities (innate, productive characteristics). In this taxonomy, skills include cognitive skills such as literacy and non-cognitive skills such as persistence and conscientiousness. The key distinction is that between attributes that are acquirable (skills) and those that are innate (abilities).

Assume that each worker potentially possesses a range of skills and can possess each of them in varying amounts. To simplify the exposition we will couch our discussion in terms of three skills. Individual earnings are determined according to some function of the skills an individual possesses and puts into use, as follows:

$$1)\ E_i = f\ (\ G_i^1,\ G_i^2,\ G_i^3\ ) + \varepsilon_i$$

3

where E_i are earnings for individual i, G_i^k is the amount of skill k that person i supplies in the market, and ε_i is a disturbance term that is independent of the skills. The disturbance term captures either individual idiosyncratic events that are independent of the skill levels or measurement error in earnings. We interpret the f(.) function as an earnings generation function derived ultimately from an overall production function that is separable in other (non-skill) inputs. Thus, by characterizing the f(.) function, we can learn about the importance of the various skills and how they interact in production. To focus ideas, we will think of G^1 as cognitive skills of the type measured in literacy tests, G^2 as other (perhaps manual) skills that are not captured in such tests, and G^3 as non-cognitive characteristics such as persistence that might be partly acquired through schooling.

Characterizing 1) would be a relatively straightforward exercise if we observed the skills, G_i^k. Typically, of course, we do not observe them. What we do observe are some of the inputs used in generating the skills. To see how they enter our framework, consider a set of production functions for generating the skills:

2) $G_i^k = h_k \ (\ yrs_i, \ \exp_i, \ \theta_i)$

where k indexes the skill type, yrs corresponds to years of formal schooling, exp is years of work experience and θ is a vector of innate abilities. The vector of abilities, θ, may include both cognitive and non-cognitive elements. That is, non-cognitive abilities such as persistence could be useful in generating both non-cognitive and cognitive skills.

If we do not observe the G_i^ks directly, we can obtain an estimating equation by substituting equation 2) into 1). This yields a reduced form specification for earnings given by:

3) $E_i = g(\ yrs_i, \ \exp_i, \ \theta_i) + \varepsilon_i$

Thus, we are considering a hierarchical model in which covariates commonly used in wage regressions are inputs into skill production and these skills (plus an independent error term) determine wages.

Now, let us examine the partial derivatives of earnings with respect to each of the skill production inputs (e.g., schooling, experience or an element of the ability vector). The partial derivative associated with one of the inputs, x, can be expressed as,

4) $\dfrac{\partial E}{\partial x} = \dfrac{\partial f}{\partial G^1} * \dfrac{\partial h_1}{\partial x} + \dfrac{\partial f}{\partial G^2} * \dfrac{\partial h_2}{\partial x} + \dfrac{\partial f}{\partial G^3} * \dfrac{\partial h_3}{\partial x}$

where we suppress the i subscript for simplicity. Thus, if x corresponds to years of schooling, then equation 4) says that the observed effect of an additional year of schooling reflects the effects of an extra

year of education on the production of each skill times the implicit price paid for that skill. It is apparent from equation 4) that with measures only of earnings and observable inputs used in producing skills, we cannot make any statements about skill production or how skills combine in production. However, if we have individual observations one or more skills, we can potentially say much more.

With G^1 observed, our quasi-reduced form earnings function becomes:

$$5)\ E_i = g*(G_i^1,\ yrs_i,\ \exp_i,\ \theta_i) + \varepsilon_i$$

Now consider the derivative of g* in equation 5) with respect to x:

$$6)\ \frac{\partial g*}{\partial x} = \frac{\partial f}{\partial G^2} * \frac{\partial h_2}{\partial x} + \frac{\partial f}{\partial G^3} * \frac{\partial h_3}{\partial x}$$

Note that we can infer the contribution of G^1 to earnings generation from the differences between the reduced form derivative 4) and the quasi-reduced form 6). Taking this difference yields:

$$7)\ \frac{\partial f}{\partial x} - \frac{\partial g*}{\partial x} = \frac{\partial f}{\partial G^1} * \frac{\partial h}{\partial x}$$

This difference (e.g., the difference between the derivative of earnings with respect to schooling first not conditioning and then conditioning on G^1) represents the fraction of the total impact of schooling on earnings due to the impact of formal schooling in generating literacy skills and impact of literacy skills on earnings. Given that we observe G^1, we can go further and derive insights into the production of G^1 through direct estimation of the skill production function 2). This is what we do in the fourth section.

Note that, as expressed in equation 4), the implicit skill prices facing an individual will be a function of the ability vector θ_i. The fact that the implicit skill price may vary with unobserved influences points to the use of quantile regressions since they effectively allow us to observe derivatives of earnings with respect to observable variables at the different values of the unobservables that generate the various conditional quantiles. Moreover, under certain conditions we could sign the relationship between the implicit skill price and unobserved ability based on how $\frac{\partial f}{\partial G^1}$ varies across increasing conditional earnings quantiles.[2] We return to this point in section 5 of the paper.

---

[2] Sufficient conditions are if (i) θ_i is a scalar rather than a vector, (ii) the h(.) functions are monotonic in θ_i and (iii) the f(.) function are monotonically increasing in skills. If so, increasing quantiles of the earnings distribution, conditional on G^1, yrs and exp, are associated with increasing values of θ_i.

3. Data

The main dataset we use in this investigation is the International Adult Literacy and Skills Survey (IALSS), the Canadian component of the Adult Literacy and Life Skills Survey (ALL). Statistics Canada carried out this survey in 2003 to study the skills of Canadians. The IALSS includes standard questions on demographics, labour force status and earnings, but it also measures literacy and related cognitive skills in four broad areas: Prose Literacy, Document Literacy, Numeracy, and Problem Solving. Perhaps of most importance for our purposes, the IALSS did not attempt to measure abilities in math and reading but tried to assess capabilities in applying skills to situations found in everyday life. Thus, the Prose questions in the surveys assess skills ranging from items such as identifying recommended dosages of aspirin from the instructions on an aspirin bottle to using "an announcement from a personnel department to answer a question that uses different phrasing from that used in the text." The Document questions, which are intended to assess capabilities to locate and use information in various forms, range from identifying percentages in categories in a pictorial graph to assessing an average price by combining several pieces of information. The Numeracy component ranges from simple addition of pieces of information on an order form to calculating the percentage of calories coming from fat in a Big Mac based on a nutritional table. Thus, the questions are related to implementation and use of skills in the real world and are intended not just to elicit current capacities but also adaptability to answering questions in other contexts (Murray, Clermont and Binkley, 2005).[3] This is an important point for the interpretation of our results since we interpret the test results as revealing job relevant skills at the time of the interview rather than inherent abilities.

In addition to providing measures of cognitive skills used in daily life, these data have three important features. First, they provide measures of skills for a representative sample of the adult population. Other measures of abilities or skills typically take the form of student achievement while in school or ability measures such as IQ taken at a young age. Second, the sample size is large, allowing analysis that would not be feasible with a much smaller sample. Our sample contains observations on 23,038 individuals, in contrast to the 1994 Canadian IALS that had 5660 observations.[4] Third, the IALSS data provide rich information on educational attainment and several key variables used in our

---

[3] The IALSS builds on the International Adult Literacy Survey (IALS) that was carried out in several countries during the period 1994 to 1998.

[4] Sample sizes for the IALS surveys carried out in the mid- to late-1990s were typically less than 6,000, even for large countries such as the U.S. and Germany. The large Canadian sample size is also unique in the current round of data collection that started in 2003.

analysis, including exact date of birth, province of birth, province in which the respondent attended secondary school, and province of current residence.

The survey covers individuals age 16 and over, and this is also the age range that we focus on in our analysis. In order to focus attention on the Canadian educational system and cognitive skill generation in Canada, we exclude from our sample anyone born outside of Canada.We also drop individuals who list their main activity as "student" in order to focus on the effect of completed schooling and what happens subsequently to individual skills and outcomes such as earnings. We also drop the over-sampled aboriginal population, reserving a careful analysis of these individuals for a separate paper. For reasons discussed later, we also focus on those with 16 years of schooling or less. The result is a sample of size 12,370, which forms the basis of our initial analysis of the determinants of cognitive skills. However, when we turn to our investigation of the impact of cognitive skills on earnings, we restrict ourselves to those employed at the time of the survey. We also drop the self-employed and workers with weekly earnings that are less than $50 and over $20,000. The latter restriction cuts out a small number of individuals with earnings that are substantial outliers relative to the rest of the sample. We exclude the self employed because we wish to assess the way skills are rewarded in the labour market, and self employed earnings reflect both that remuneration and returns to capital. We include both males and females throughout, dividing the analysis on gender lines in some places. Finally, we use the sample weights throughout the analysis, so all summary statistics and regression estimates are nationally representative.

For the earnings analysis our dependent variable is weekly earnings. In the IALSS respondents are first asked about their standard pay period and then asked about typical earnings in that pay period. Using these responses we construct a weekly earnings measure for each paid worker. Thus, for example, in the case of individuals who are paid monthly we divide their usual monthly earnings by 4.333.

A salient feature of the data is the strong correlation among the various cognitive skill measures. The correlation between the Prose literacy and Document literacy scores is 0.96, that between Prose literacy and Numeracy is 0.90, and the correlation between Prose literacy and Problem Solving is 0.93. Further, a principal components analysis indicated only two principal components with the first being vastly more important and placing equal weight on all four scores. Thus a simple average of the four scores captures much of the information available in the skill measures. This is the skill measure that we use in much of the analysis, but we do report some results for the four individual skills.

We also compiled data on compulsory schooling laws in Canada. We use the minimum school leaving age as an instrument for schooling.[5] Changes in compulsory schooling laws have been shown to have significant effects on educational attainment, and have been a commonly used instrument for education in many studies (see, for example, Acemoglu and Angrist, 2000; Lochner and Moretti, 2004; Milligan, Moretti and Oreopoulos, 2004; and Oreopoulos, 2006). In Canada, education falls under provincial jurisdiction and compulsory schooling laws vary across provinces as well as over time within provinces. Based on detailed information on compulsory schooling in Canadian provinces we created three indicator variables for the school leaving age: <15, 15 to <16, and 16 or more, as explained more fully below.[6] Likewise, we created three indicator variables for the school entry age: <7, 7 and 8 years.

Table 1 reports summary statistics for our sample. The literacy test outcomes are scaled to fall between 0 and 500, with our average of the four scores having a mean of 261 and a wide range. Females are slightly over-represented in the sample relative to the population but we control for gender in all of our estimates. The average number of years of completed schooling (11.5) is less than what is found in other Canadian surveys, reflecting the exclusion of those with more than 16 years of schooling. Comparing this number to the fact that over 50% of respondents report that their parents were high school dropouts reflects strong progress in education across generations. What we call the "ability proxies" (the variables containing recollections of perceptions of school) exhibit real variability, implying they may be useful as covariates.

The minimum school leaving age has increased over time in all provinces, and the maximum school entry age has declined. However, at any point in time there is considerable variation across jurisdictions in compulsory schooling provisions. About one-third of the sample faced a school leaving age of 15, and almost half the sample faced school leaving ages of 16 years or higher. Close to one-half of the sample faced a school entry age of less than 7 years.

4. The Generation of Cognitive Skills

This section examines the sources of literacy, numeracy and problem-solving skills (which we refer to simply as "cognitive skills" or "literacy skills"). We focus on the average skill score as our representative measure. Our regressions use the log of the average score as the dependent variable so our estimated coefficients show impacts in terms of percentage changes in skills.

---

[5] We compiled data on the school leaving age from provincial statutes and the date of proclamation or ascension from provincial regulations. Our data are similar to those compiled by Oreopoulos (2006).
[6] When there are exemptions or other requirements, we use the main and most basic school leaving age. We also use the urban school leaving age if both urban and rural ages are specified.

Before presenting the estimation results, we set out a brief, heuristic model of cognitive skill generation. The model will help to put our estimates in context as well as providing guidance in thinking about identification issues. Consider a simple model in which individuals start out at birth endowed with two key characteristics: their ability and parental resources. By parental resources, we mean something quite broad, incorporating both parental income and parental willingness and ability to support their children's education and literacy acquisition. Pre-school children begin to acquire literacy based on these fundamental characteristics. Once they enter school, these characteristics interact with characteristics of the education system such as teacher quality, class size and the attitudes and abilities of peers. New additions to cognitive skills with each year of schooling are then functions of ability, parental resources, school characteristics and the literacy and numeracy level at the beginning of the period. These influences may interact in complicated ways. These additions continue until the legal school leaving age. After that point until the end of high school, students make a decision each year on whether to continue in school. That decision will again be a function of ability, parental resources and school characteristics, but it is also likely to depend on literacy acquired to that point. The more literate and numerate a student is, the less onerous they are likely to find school and, thus, the more likely they are to choose to stay an extra year. Finally, after high school, whether an individual continues to go to school will be determined by a combination of their own decision to apply to continue and the decision of the college or university on whether to admit them. The latter decision will likely be a function of the student's cognitive skills as reflected in her grades. Thus, schooling and cognitive skills are co-determined with extra years of schooling leading to increased literacy and numeracy but increased skills also leading to more years of schooling, especially after the legal school leaving age. Indeed, once we account for expectations, the inter-relation between the two may be even tighter. Individuals who do not expect to continue with school past the legal minimum may rationally under-invest in acquiring literacy and numeracy skills while they are in school.

Once individuals leave school, skill acquisition is likely more difficult. Literacy and numeracy skills may be acquired on the job if they are needed for carrying out tasks at work but otherwise further acquisition would require active investment in non-work hours. Indeed, it seems quite possible that individuals could lose cognitive skills after they leave formal schooling if those skills depreciate when they are not used.

We are interested in characterizing as many of the components of literacy, numeracy and problem-solving generation as possible. In particular, we are interested in the relationship of literacy and

numeracy to parental resources since that relationship is fundamentally linked to the extent that one generation's literacy hinges on the resources of the previous generation. We are also interested in the relationship between formal schooling and cognitive skills since this is a major channel through which we could hope to affect the skill distribution. Finally, we are interested in whether literacy and numeracy decline or rise after leaving school -- whether cognitive skills have a "use it or lose it" character -- and how this process is related to characteristics such as age or work experience.

Figure 1 shows the partial relationship between cognitive skills and years of schooling, after controlling for age, gender, parental characteristics, province of residence and urban/rural residence. The relationship is upward sloping, close to linear, but with a small amount of concavity, especially after 16 years of schooling. The latter probably reflects the fact that most of the skills assessed by the IALSS tests are likely to be acquired at the elementary and secondary levels rather than at the post-secondary level. In the regression results that follow we restrict the sample to those with 16 years of schooling or less. Within this range the partial relationship between average skills and years of schooling is approximately linear, which simplifies our empirical analysis.[7] This is also the range within which compulsory schooling laws, one of our instrumental variables for education, are most likely to be binding.

The first column of Table 2a presents our simplest OLS regression in which the dependent variable is the log of the average skill score and the independent variables are age, age squared, years of schooling, gender, dummies for residence in small and large urban areas, and province of residence.[8] All the variables are statistically significant but this does not mean their actual impacts are sizeable. Thus, the estimates show that women have lower average skills than men (conditional on schooling and age) but only by 1.2%. Similarly, the age and age-squared coefficients are highly statistically significant but together they imply that the impact of an extra year of age on skills is actually -0.1% at age 30. This finding that there is a relatively weak relationship between literacy and either age (or experience) is a key part of the discussion in Green and Riddell (2003). The one relationship that is economically substantial is the one between cognitive skills and schooling. One extra year of schooling increases literacy and numeracy by 3.6%.

---

[7] When we restrict the sample to those with 16 years of schooling or less, an OLS regression of log average skill on years of schooling, years of schooling squared and other covariates listed in Table 2 yields an insignificant coefficient on the quadratic years of schooling term.

[8] Estimated coefficients on the urban – rural and province of residence dummies are not shown in the table.

As we discussed earlier, literacy and years of schooling are likely to be jointly determined. In that case, OLS estimates are likely to be biased. We attempt to address this in two ways. First, biases may arise because of a correlation between literacy and schooling arising from unobserved variables that are correlated with both education and literacy. One important set of variables that is often not available consists of parental and family background characteristics. Here we fortunately are able to control for a variety of family background factors including educational attainment, occupation and immigrant status of the respondent's mother and father. Other potential unobserved factors include individual characteristics such as innate ability and motivation. If high ability people do not view it as particularly costly to either acquire literacy or go to school then we could observe a positive coefficient on schooling because years of schooling is proxying for ability rather than as a reflection of a causal impact of schooling on literacy. This problem can be addressed if we have a measure of ability since once we control for ability, any relationship between schooling and literacy cannot be due to an omitted ability term. Note, though, that many studies that try to control for ability (in, for example, earnings regressions) actually use scores on tests much like our literacy tests. What we would require is a test score from a very young age - before the process we are trying to study really begins. Since we don't have that, we instead try to proxy for ability using two variables that are plausibly related to it.

In the second column of Table 2 we add variables on parental education and parental immigrant status. Introducing these variables has virtually no impact on the gender coefficient, but it does have the expected consequence of reducing the coefficient on years of schooling, by about 11%. Including them also leads to an increase in the age coefficient. Given that the coefficient on the age-squared variable also becomes more negative, the net effect of age remains quite small. The parental education variables are jointly highly significantly different from zero but, perhaps surprising, the effect is found mainly at low levels of parental education. Having a parent (either mother or father) who is a high school drop out decreases average literacy by 3% to 4%. However, parental education beyond high school graduation has relatively modest further impacts on literacy. Interestingly, not knowing a parent's education level (which is the case for approximately 9% of the sample) has a strong effect, being associated with approximately 5%-7% lower literacy. While we included this variable in order to allow us keep the observations with missing information on parental education, it seems possible it represents something real. For example, children who do not know a parent's education likely did not have a close relationship with that parent. Thus, the estimated coefficient may reflect the extent to which literacy is generated through direct parental involvement. Finally, having a mother who is an immigrant has no

association with literacy while the coefficient on the immigrant father dummy is positive but small. We also tested specifications in which we included a set of parental occupation dummy variables but these were never jointly statistically significant.[9] We also find that a dummy variable representing whether the individual's mother was working when the individual was 16 does not have a statistically significant effect. Overall, the results point to a surprisingly weak association between literacy and parental background once we control for the individual's education. Only schooling seems to have a substantial impact on literacy generation.

In the third regression, we add the ability proxies. Both are based on the respondent's experiences while in secondary school. The first is a dummy variable equalling one if the person agreed or strongly agreed with the statement that they got good grades in math when they were in school and the second is a dummy variable equalling one if the respondent agreed or strongly agreed with the statement that teachers often went too fast and the person often got lost. Either of these could plausibly be seen as proxies for innate ability. Both of these variables enter significantly, with people who claimed to have gotten good grades in math having 2.5% higher literacy and those who thought teachers went too fast having 2.6% lower literacy. Including these variables results in the gender coefficient becoming small and statistically insignificant, but has no impact on the age coefficients. Their inclusion does reduce the estimated impact of schooling on literacy, but the magnitude of the decline is modest (about 10%).

An alternative approach to the problem of identifying a causal effect is to use an instrumental variable strategy. We use compulsory schooling laws as instruments for education. Both the maximum school entry age and the minimum school leaving age vary over time and across provinces. Changes in these laws have been shown to have significant effects on educational attainment, and have often been used as an instrument for education. Based on detailed information on compulsory schooling in Canadian provinces we created three indicator variables for the school entry age (<7, 7, and 8) and three indicators for the school leaving age: <15, 15 to <16 and 16 or more. Because we observe the exact date of birth as well as the province in which the respondent attended secondary school, we can match individuals more precisely than is done in previous studies to the minimum school leaving age that they

---

[9] In particular, a test of the hypothesis that the set of father's occupation dummy variables jointly had zero effects has an associated P-value of .13. The same test for mother's occupation has a P-value of .79.

faced while in high school.[10] Similarly, because we observe province of birth we can match individuals more precisely than in many previous studies to the school entry age provisions that they faced at the school entry stage. Furthermore, our school entry and school leaving age data take account of the fact that changes in these laws are often implemented during a calendar year (e.g. July 1 or September 1) and may have differential effects on individuals born in the same year, depending on their date of birth. Identification of the causal impacts of education on literacy skills assumes that variations in the minimum school leaving age over time and across provinces induced changes in years of schooling that are unrelated to unobserved factors such as ability or motivation.

Although we have a suitable instrument for years of schooling, we were unable to obtain a separate instrument for schooling squared. A standard approach to instrumenting for a squared term would be to use higher order terms in the instrument as instruments for the squared term. However, since our instrument consists of a set of dummy variables, the higher order terms equal the dummy variable itself and we are left without a separate instrument for the squared term. We therefore restrict the analysis to those with 16 years of school or less, the sub-sample for which the partial relationship between log literacy and years of education is approximately linear. This is also the range of schooling within which compulsory schooling laws are likely to be binding.

The results from our two stage least squares estimation using these instrumental variables are reported in columns 4 (which uses the SLA as instruments) and 5 (which includes both the SLA and SEA as IVs) of Table 2a. The first stage regressions, in which years of schooling is the dependent variable, are reported in Table 3. These indicate that, as expected, parental education is strongly positively related to years of schooling. There are also some interesting gender differences. For example, having a father with a university bachelor's degree has a large and statistically significant coefficient while having a mother with a university degree does not. Having an immigrant mother and living in a major urban area are also positively associated with years of schooling.

Both sets of instruments are jointly statistically significant, indicating that the instrument exerts a significant influence on the endogenous variable is satisfied. The omitted category for the SLA is <15, and that for the SEA is 8 years of age, so the coefficients on the included compulsory schooling law variables are expected to be positive. The first stage F statistics corresponding to the hypothesis that the instrumental variables are jointly equal to zero are 15.8 and 10.0 for the school leaving age and SLA

---

[10] Most previous studies match individuals based on their province or state of birth and the year in which they turned 14. This introduces measurement error for two reasons: (i) some individuals change location between birth and age 14, (ii) compulsory schooling laws can affect individuals born in the same year differently, depending on their month of birth.

plus SEA instruments respectively. The former is comfortably above the "rule of thumb" critical value for weak instruments (a first stage F-statistic = 10) while the latter equals that critical value. Thus we place most emphasis on results based on the school leaving age instrument. The estimated coefficients associated with the school leaving age variables indicate that raising the school leaving age to 15 is associated with an increase in years of schooling of approximately 0.6 years on average, and an increase in the minimum dropout age to 16 also increases educational attainment by 0.6 years.

The IV estimates in columns 4 and 5 of Table 2a are broadly similar. We will focus on the estimates associated with the school leaving law instrument. The IV estimate of the schooling effect equals 0.73, more than double the corresponding estimate in column 2, implying even stronger schooling effects than those estimated with OLS. Interestingly, once we instrument for schooling, the parental background variables become much smaller in magnitude, and none of these are statistically significant. The age coefficients also decline in magnitude and are no longer statistically significant. Apart from formal schooling, only gender exerts a significant impact on literacy. We do not present the coefficients corresponding to the provincial dummy variables for the sake of brevity but they show that the Atlantic provinces and Ontario have essentially similar cognitive skill levels, Quebec has significantly lower skill levels and the Prairies and BC all have significantly higher levels. Our main conclusion is that, if the assumptions underlying our instruments are correct, these results indicate that education has a strong causal effect on cognitive skills and that schooling is the dominant determinant of literacy, numeracy and problem-solving skills. To put the estimated effects in perspective, completing four extra years of schooling (e.g., moving from being a high school graduate to a university graduate) implies a 29% increase in literacy, based on the school leaving age IV estimates. This would be enough to move the individual from the median to above the 80th percentile of the cognitive skill distribution in 2003.

We also estimated the specifications in columns 4 and 5 by limited information maximum likelihood (LIML). As is well known, two stage least squares estimates are consistent (if the assumptions regarding the instruments are valid) but are biased in small samples. Bias is especially likely if the instruments are weak and there are many over-identifying restrictions. LIML estimates are also consistent but have the potential advantage of providing a small sample correction for the bias in two stage least squares. The LIML estimates (and associated standard errors) of the coefficient on years of schooling based on the two sets of instrumental variables are as follows:

LIML Estimates- Literacy

Variable	OLS2	IV1	LIML1	IV2	LIML2
Years of Schooling	0.031***	0.073***	0.075***	0.084***	0.096***
	[0.001]	[0.012]	[0.012]	[0.012]	[0.017]
Observations	12370	12370	12370	12370	12370

The TSLS and LIML estimates are very similar and are not significantly different from each other. Furthermore, there is no evidence that the TSLS estimates are biased upwards in our sample. Indeed, if anything the LIML estimates are slightly larger in size than their TSLS counterparts.

Table 2b reports IV estimates of log skills equations for each of the individual skills, using compulsory schooling laws as instruments for education. The specification is thus the same as that in column 4 of Table 2a. Although the results are broadly similar to those based on the average skill score, there are some noteworthy differences in the coefficients on gender and years of schooling. The Table 2a estimates indicate the women have average skills that are 2.4% below those of men after controlling for other influences. This small gender gap in average skills reflects the offsetting effects of lower document literacy and substantially lower numeracy skills among women but prose literacy and problem solving skills that do not differ significantly from those of men. The estimated impacts of formal schooling on skills vary little across the four individual skills, although the estimated coefficient on document literacy is somewhat larger than that on the other three skills.

5. Cognitive Skills and the Returns to Schooling

The results in the previous section provide strong evidence that formal education exerts a causal influence on basic literacy and numeracy skills. We now turn to the second major objective of this paper – to estimate the extent to which the estimated returns to schooling reflect the combined impact of education on literacy and numeracy skills and the payoff to these skills in the labour market. To do so we empirically implement the framework described in section 2 using our earners sample, which consists of those employed at the time of the survey. This sample is younger and more highly educated than the full sample used in the previous analysis. Being younger and more educated, they may be less influenced by changes in compulsory schooling laws.

Figure 2 plots the partial relationship between years of schooling and log weekly earnings for the full earners sample. The partial relationship is approximately linear over the range 10 to 17 years of schooling, and exhibits some evidence of diminishing returns past 17 years of education. There is also evidence of above-average earnings gains associated with achieving 12, 14, 16 and 17 years of education, which could reflect "sheepskin effects" associated with high school graduation, college completion, and completion of university programs at the bachelor's and master's level.

We present estimation results from mean regressions using the log of weekly wages in Table 4. The first column shows the results from a standard regression with a female dummy, years of schooling, experience and experience squared as covariates. As before, controls for parental characteristics, province of residence and urban/rural status are included but not reported. The results are extremely standard in terms of their magnitudes and sign patterns (see Card (1999) for a review of the large relevant literature). In the second column, we add the average cognitive skill variable. We have also estimated specifications in which we include all 4 scores separately. In those estimations, document literacy enters statistically significantly with a coefficient of .0021, numeracy enters statistically significantly with a coefficient of .0011 and problem solving and prose literacy have smaller, not statistically significant and offsetting coefficients. Note that these significant separate effects essentially add up to the estimated coefficient on the average skill measure presented in Table 4. This suggests that numeracy may have separate effects from the other three types of literacy and that its effects are smaller than whatever is being captured (primarily) in the document score.

Adding the average skill score leads to a reduction in the derivative of log earnings with respect to education from .085 without the skills variable to .070 when it is included. This is a reduction of about 18%, suggesting both that literacy and numeracy skills play an important role in the returns to education and that education has a substantial impact on earnings over and above the impact related to production of literacy and numeracy skills. In contrast to the effect on the schooling coefficient, the coefficients on the experience variables are unchanged when we introduce the cognitive skills variable. This is a direct reflection of the fact that literacy and numeracy generation is not related to age or experience in the cross-section. In terms of the framework set out above, experience does not enter the skill production function and so the first term on the right hand side of equation 4 is zero. The implication is then that the derivative with respect to experience is the same whether or not we condition on cognitive skills. Finally, the direct impact of cognitive skills on earnings is substantial. A 25-point

increase in the average skill score (the equivalent of about 1/2 of a standard deviation in the skill score distribution) has an impact equivalent to an extra year of schooling.

As mentioned in our theoretical discussion, our estimation may be affected by omitted variables bias. In particular, the error term in the regression may include various types of ability that are correlated with the included variables. Typically, ability is assumed to affect both schooling choices and earnings, leading to biased estimates. Given our specified model, if we assume that unobserved cognitive abilities only affect the generation of cognitive skills and other, non-cognitive abilities do not affect the generation of cognitive skills then literacy will not be correlated with the error term and does not, itself, represent an endogeneity problem. However, even in that case we still need to address the potential endogeneity of schooling. Under less restrictive assumptions about the role of unobserved factors, we need to address the potential endogenity of both education and cognitive skills. We do so using two sets of instruments. The first is the SLA, as employed previously in the cognitive skills estimation.

The second set of instrumental variables is the province where the individual resided when he or she was last in high school or middle school fully interacted with age. The idea behind this instrument, which is similar to that used by Card and Krueger (1992) in their analysis of school quality in the U.S., is that different levels of public resources applied to schooling in different provinces for different generations will lead to different schooling attainment for otherwise identical individuals. This instrument will be valid if provincial education resources and policies while in high school influence schooling outcomes but do not directly influence the production of cognitive skills. In implementing this approach it is important to control for current province of residence in both the first stage (schooling regression) and the second stage (cognitive skills regression). Province of current residence may be related to skills if more literate individuals choose to migrate to provinces with a higher proportion of high skill jobs and low literacy individuals chose to move to, for example, provinces with large numbers of resource sector jobs. In that case, to the extent that province of residence during high school and current province of residence are correlated, the province of residence would pick up this migration effect rather than the schooling effect we want it to capture. Controlling for province of current residence addresses this problem and means that we are identifying the schooling effect from people who currently reside in the same province but were schooled in different provinces at different times.

The results from instrumenting for schooling when the cognitive skills variable is not included are reported in the third and fifth columns of Table 4. Estimation by instrumental variables has very little effect on the coefficient on schooling, which is slightly larger with the school leaving age IV and almost

17

identical in size under estimation with the province of high school IV. In the fifth column, we repeat this exercise but also include the average skill measure. Again, instrumenting yields schooling coefficients that are somewhat larger for the compulsory schooling instrument and very similar in size to the simple OLS results in column 2 for the province of high school instrument. Comparing the estimates in columns 3 and 4, as well as those in columns 5 and 6, introducing the average skill score reduces the schooling coefficient by approximately 40% for the school leaving age IV and 35% for the province of high school IV. The skill score coefficient estimated by TSLS is about 60% larger in size than the OLS coefficient in column 2. Thus, the IV estimates imply that cognitive skill generation accounts for about 35% to 40% of the returns to schooling estimated in standard models of earnings.

We also estimated the specifications in columns 3 to 5 of Table 4 by LIML. The estimates of the coefficients on years of schooling and literacy, shown below, are similar to their TSLS counterparts:

LIML Estimates- Earnings

Variable	IV2	LIML2	IV3		LIML3	
Years of Schooling	0.089***	0.089***	0.084***	0.054***	0.084***	0.031
	[0.026]	[0.028]	[0.005]	[0.015]	[0.005]	[0.039]
Average Literacy Score				0.005**		0.009
				[0.003]		[0.007]
Observations	7569	7569	7569	7569	7569	7569

As discussed earlier, our theoretical framework points to advantages from using a quantile regression framework. We present the results of quantile regression estimation for the 10th, 25th, 50th, 75th and 90th quantiles in Table 5. The key implications from the estimation are as follows. First, returns to both schooling and experience decline across quantiles. The finding of heterogeneity in returns to education across the earnings distribution has been observed by previous authors. Buchinsky(1997) finds returns to education that rise across quantiles for all experience groups. Aris et al. (2001) estimate similar quantile regressions using US twins data and incorporating approaches to address endogeneity. With non-IV estimation, they find that the coefficient on education rises from the 10th to the 50th percentile but does not change across the upper portion of the distribution. When using instruments to

address measurement error and twins status to address ability bias, their estimated schooling coefficients appear relatively similar across the distribution but are not very precisely estimated in the tails.

Perhaps the most interesting result in Table 5 is the relative lack of variation in the coefficient on the cognitive skills measure across the quantiles. While the coefficient for the 10th quantile appears substantively smaller than those at the other quantiles it is not actually statistically significantly so. Moreover, if we run quantile regressions at the quantiles directly surrounding the 10th quantile (e.g., the 5th and 15th quantiles), we obtain estimated literacy effects that are almost exactly the same as those reported for the upper quantiles - the 10th quantile appears to be a bit of a statistical outlier. In the context of our theoretical model this implies that cognitive skills do not interact with other attributes in earnings generation. Thus, other attributes or skills such as persistence and leadership skills (Kuhn and Weinberger (2005)) may contribute to individual earnings but their contributions are not enhanced by having more literacy skills. Basic cognitive skills are not a silver bullet that both contributes directly to earnings and increases returns to other attributes.

In summary, our results suggest that literacy, numeracy and problem-solving skills are important determinants of earnings but that there is a great deal of earnings variation that is accounted for by other factors. According to our IV estimates, 35% to 40% of the return to schooling estimated in previous studies can be attributed to the combined effect of formal education on these skills and the value placed on literacy and numeracy in the labour market.

6. Conclusions

A substantial body of recent research provides evidence that formal education exerts a powerful influence on individuals' lifetime earnings. This research concludes that this influence is causal in nature, rather than simply reflecting a positive correlation between schooling and earnings. In addition there is growing evidence of causal impacts of education on various non-pecuniary outcomes, such as civic participation, health and longevity, participation in crime, and life satisfaction. However, little is known about the mechanisms that underlie these causal impacts. This paper investigates the extent to which the estimated impacts of formal schooling on outcomes reflects the impact of education on the production of literacy, numeracy and problem solving skills and the influence of these skills on individual outcomes such as earnings. To do so we use a rich data set containing measures of literacy, numeracy and problem-solving skills for a representative sample of the adult population.

Our investigation yields several noteworthy findings. First, we provide strong evidence that education has a substantial causal effect on cognitive skills, and that formal schooling is the dominant

determinant of basic literacy, numeracy and problem-solving skills. Our instrumental variables estimates indicate that each additional year of schooling raises average skills by 7% or about one-eighth of a standard deviation of the skill score distribution. To put this in perspective, completing four additional years of schooling (e.g. moving from being a high school graduate to a university graduate) would move the individual from the median to above the 80th percentile in the skill distribution. In addition, age (or work experience) has little impact on literacy and numeracy skills, suggesting that the positive relationship between experience and earnings arises for other reasons. Furthermore, parental characteristics have only modest effects on literacy and numeracy skills. Having a mother or father with less than high school education has a modest negative effect on cognitive skills, but having a parent with education beyond secondary school has no effect. Similarly, having immigrant parents has no effect. The influence of parental characteristics on skills arises indirectly through their powerful influence on the child's education rather than directly.

When we estimate earnings equations similar to those that have appeared in previous papers we obtain results similar to those in the literature on schooling and earnings: returns to schooling in the order of 7% to 8%. However, when we also control for cognitive skills these coefficient estimates drop by 35% to 40%. The difference between these two sets of estimates represents the combined effect of education on the production of cognitive skills and the value placed on these skills in the labour market. Thus according to our estimates, about one-third of the return to schooling represents the role of formal education in the production of literacy, numeracy and problem-solving skills. This is a substantial component of the return to education, but the fact that the coefficient on schooling remains large and statistically significant after controlling for skills suggests that schooling also affects earnings via other mechanisms. We also find that the direct effect of cognitive skills on earnings is substantial. A 25-point increase in literacy and numeracy skills (half of a standard deviation) is associated with an increase in earnings equivalent to an additional year of schooling.

The conclusion that formal schooling exerts a powerful causal impact on literacy and numeracy skills, together with the evidence that these skills earn a substantial return in the labour market supports the human capital perspective on the relationship between education and earnings, in contrast to the signalling view. Although such influences are widely believed to exist, to our knowledge this is the first

study to provide evidence of this causal linkage, at least in the case of literacy, numeracy and problem-solving skills of the adult population.[11]

---

[11] Cascio and Lewis (2006) provide evidence that schooling exerts a causal influence on the AFQT score of young adults. Carllson, Dahl, Ockert and Rooth (forthcoming) find that schooling raises cognitive test scores on crystallized intelligence tests among 18-year old Swedish males but has no impact on fluid intelligence test scores.

References

Acemoglu, Daron and Joshua Angrist. "How Large are Human Capital Externalities? Evidence from Compulsory Schooling Laws" *NBER Macroeconomics Annual 2000*, 2001, 9-59.

Aris, O., K.F. Hallock and W. Sosa-Escudero. "Individual Heterogeneity in the Returns to Schooling: Instrumental Variables Quantile Regression Using Twins Data," *Empirical Economics* 26 (2001) 7-40.

Black, Sandra, Paul Devereux and Kjell Salvanes. "Why the Apple Doesn't Fall Far: Understanding Intergenerational Transmission of Human Capital" *American Economic Review* 95 (March 2005) 437-449.

Buchinsky, M. "Recent Advances in Quantile Regression Models: A Practical Guideline for Empirical Research," *The Journal of Human Resources*, 33 (No.1, 1997) 88-126.

Card, David. "The Causal Effect of Education on Earnings" in *Handbook of Labor Economics* volume 3A, edited by Orley Ashenfelter and David Card. Amsterdam: North Holland, 1999, pp. 1801-1863.

Card, David. "Estimating the Return to Schooling: Progress on Some Persistent Econometric Problems" *Econometrica* 69 (September 2001) 1127-1160.

Card, David and Alan B. Krueger. "Does School Quality Matter: Returns to Education and the Characteristics of Public Schools in the U.S." *Journal of Political Economy* 100 (February 1992).

Carllson, Magnus, Gordon Dahl, Bjorn Ockert and Dan-Olof Rooth. "The effect of schooling on cognitive skills." *Review of Economics and Statistics*, forthcoming.

Cascio, Elizabeth and Ethan Lewis. "Schooling and the AFQT" *Journal of Human Resources* 41 (Spring 2006) 294-318.

Clark, Damon and Heather Royer. "The Effect of Education on Adult Mortality and Health: Evidence from Britain" *American Economic Review* 103 (6, 2013) 2087-2120.

Dee, Thomas S. "Are There Civic Returns to Education?" *Journal of Public Economics* 88 (August 2004) pp. 1696-1712.

Green , David A. and W. Craig Riddell. "Literacy and Earnings: An Investigation of the Interaction of Cognitive and Unobserved Skills in Earnings Generation." *Labour Economics* 10 (April 2003) 165-185

Grossman, Michael. "Education and non-market outcomes" Chapter 10 in *Handbook of the Economics of Education*, volume 1. Edited by Eric Hanushek and Finis Welch. New York: Elsevier, 2006, pp. 578-635.

Grossman, Michael. "The Relationship between Health and Schooling: What's New? IZA Discussion Paper 9369, September 2015.

Kuhn, P. and C. Weinberger. "Leadership Skills and Labor Market Outcomes," *Journal of Labor Economics*, 23 (July, 2005) 395-436.

Lleras-Muney, Adriana "The Relationship Between Education and Adult Mortality in the United States" *Review of Economic Studies* 72 (No. 1, 2005) 189-221.

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

Milligan, Kevin, Enrico Moretti and Philip Oreopoulos. "Does Education Improve Citizenship? Evidence from the U.S. and the U.K." *Journal of Public Economics* 88 (August 2004) 1667-1695.

Murray, T. Scott, Yvan Clermont and Marilyn Binkley. *Measuring Adult Literacy and Life Skills: New Frameworks for Assessment*. Ottawa: Statistics Canada, 2005.

Oreopoulos, Philip. 2006. "The compelling effects of compulsory schooling: Evidence from Canada" *Canadian Journal of Economics* 39 (February) 22-52.

Oreopoulos, Philip. "Do dropouts drop out too soon? Wealth, health and happiness from compulsory schooling" *Journal of Public Economics* 91 (2007) 2213-2229.

Oreopoulos, Philip, Marianne Page and Ann Stevens. "Does Human Capital Transfer from Parent to Child? The Intergenerational Effects of Compulsory Schooling" *Journal of Labor Economics* 24 (October 2006) 729-760.

Oreopoulos, Philip, and Kjell Salvanes. 2011. "Priceless: The Nonpecuniary Benefits of Schooling" *Journal of Economic Perspectives* 25 (Winter): 159-184.

Plug, Erik. "Estimating the Effect of Mother's Schooling on Children's Schooling Using a Sample of Adoptees" *American Economic Review* 94 (March 2004) 358-368.

Rosen, Sherwin. 1974. "Hedonic Prices and Implicit Markets: Product Differentiation in Pure Competition" *Journal of Political Economy*

Sen, A. *Development as Freedom*. New York: Anchor Books, 1999.

Welch, Finis. 1969. "Linear Synthesis of Skill Distributions." *Journal of Human Resources* (Summer) 311-329.

Figure 1

Regression-Adjusted log skill by Years of Schooling with School 0-8 as Base Category

Figure 2

Regression-Adjusted log weekly earning by Years of Schooling with School 0-8 as Base Category
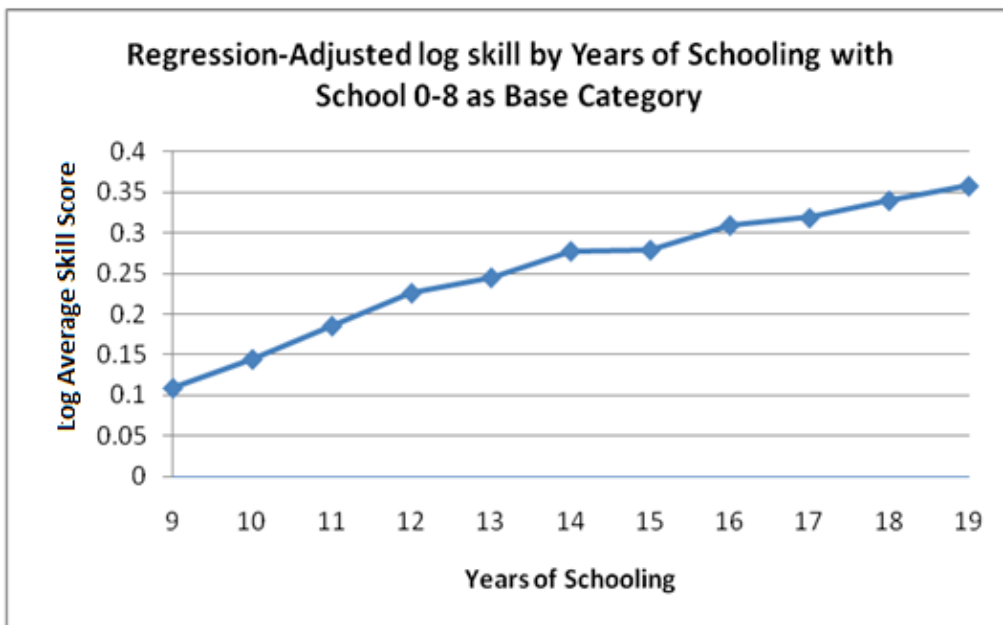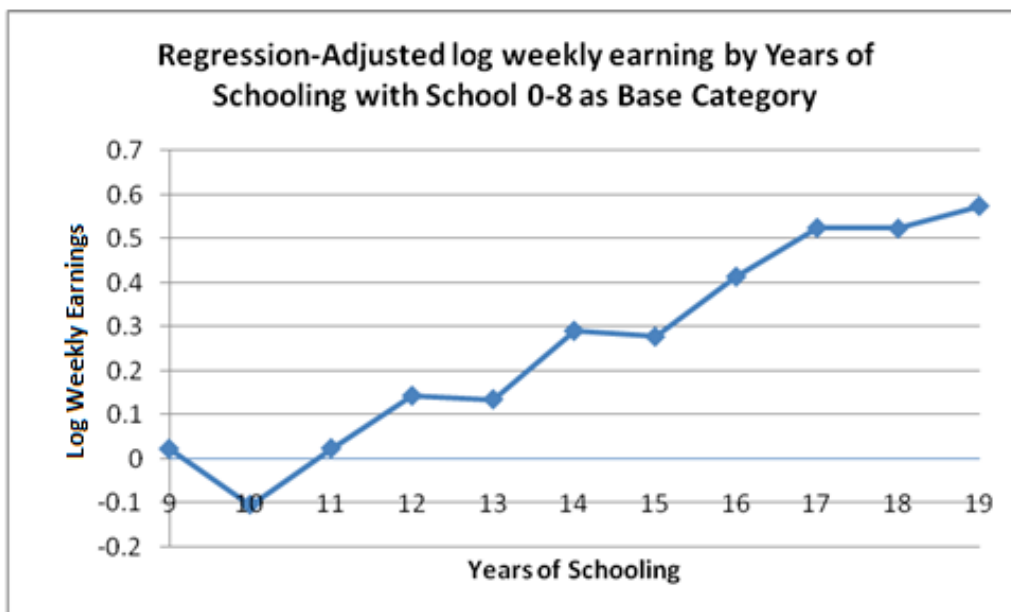
Table 1: Summary Statistics

Variables	Mean	Std. Dev.	Min	Max
DEPENDENT VARIABLES				
Prose Literacy Score	266.82	51.27	83.80	435.80
Document Literacy Score	263.52	53.12	82.80	430.20
Numeracy Score	254.33	52.99	72.00	429.60
Problem Solving Score	259.70	48.29	35.00	408.20
Average Literacy Score	261.09	49.74	96.95	416.05
EXPLANATORY VARIABLES				
Female	0.545	0.498	0	1
Years of Schooling	11.577	2.807	0	16
Age of Respondent				
Age	47.17	17.24	16	100
Age: 26-35	0.149	0.356	0	1
Age: 36-45	0.243	0.429	0	1
Age: 46-55	0.246	0.431	0	1
Age: 56-65	0.176	0.381	0	1
Age: 66+	0.186	0.389	0	1
Mother's Education				
Less than High School	0.544	0.498	0	1
High School	0.205	0.404	0	1
Some Post Secondary	0.108	0.311	0	1
University Degree	0.050	0.218	0	1
None Reported	0.092	0.290	0	1
Father's Education				
Less than High School	0.584	0.493	0	1
High School	0.150	0.357	0	1
Some Post Secondary	0.098	0.297	0	1
University Degree	0.065	0.246	0	1
None Reported	0.104	0.306	0	1
Parental Immigration				
Immigrant Mother	0.094	0.292	0	1
Immigrant Father	0.109	0.311	0	1
Ability Proxies				
Good Math Grades	0.716	0.451	0	1

Teachers Too Fast	0.300	0.458	0	1
School Entry Age				
<7 years	0.485	0.500	0	1
7 years	0.386	0.487	0	1
8 years	0.129	0.336	0	1
School Leaving Age				
<15 years	0.205	0.404	0	1
15 years	0.278	0.448	0	1
16 years	0.517	0.500	0	1

Table 2a: Log of Average Skill Score Regressions

Variable	OLS1	OLS2	OLS3	IV1	IV2
Years of Schooling	0.036***	0.031***	0.028***	0.073***	0.084***
	[0.001]	[0.001]	[0.001]	[0.012]	[0.012]
Female	-0.012**	-0.010**	-0.007	-0.024***	-0.027***
	[0.005]	[0.005]	[0.005]	[0.007]	[0.007]
Age	0.003***	0.006***	0.006***	-0.001	-0.003
	[0.001]	[0.001]	[0.001]	[0.002]	[0.002]
Age Squared	-0.007***	-0.009***	-0.009***	0	0.002
	[0.001]	[0.001]	[0.001]	[0.003]	[0.003]
Mother's Education					
Less than High School		-0.040***	-0.043***	-0.008	0.001
		[0.006]	[0.006]	[0.012]	[0.013]
Some Post Secondary		0	-0.002	-0.015	-0.019*
		[0.007]	[0.007]	[0.010]	[0.011]
BA or More		0.012	0.01	0.007	0.005
		[0.012]	[0.011]	[0.015]	[0.016]
None Reported		-0.066***	-0.068***	-0.011	0.004
		[0.011]	[0.011]	[0.019]	[0.021]
Father's Education					
Less than High School		-0.026***	-0.024***	0.008	0.017
		[0.007]	[0.007]	[0.013]	[0.014]
Some Post Secondary		0.002	0.004	-0.007	-0.009
		[0.008]	[0.008]	[0.010]	[0.011]
BA or More		0.018*	0.024***	-0.005	-0.011
		[0.009]	[0.009]	[0.014]	[0.015]
None Reported		-0.047***	-0.045***	0.005	0.019
		[0.011]	[0.011]	[0.019]	[0.020]
Immigrant Mother		0.006	0.005	-0.013	-0.019
		[0.008]	[0.008]	[0.011]	[0.012]
Immigrant Father		0.016**	0.018**	0.004	0
		[0.008]	[0.007]	[0.010]	[0.011]

Good Math Grades			0.025*** [0.006]		
Teachers Too Fast			-0.026*** [0.006]		
Constant	5.185*** [0.021]	5.190*** [0.021]	5.221*** [0.022]	4.788*** [0.116]	4.680*** [0.122]
Observations	12370	12370	12370	12370	12370
R-squared	0.47	0.50	0.51	0.27	0.13

Notes:
1. Standard errors in parentheses. *, **, *** statistically significant at 10%, 5%, 1% level.

Table 2b: Individual Skill Score Regressions with CSLA Instrument

Variable	Log Prose Score	Log Document Score	Log Numeracy Score	Log Problem Solving Score
Years of Schooling	0.070***	0.085***	0.070***	0.069***
	[0.012]	[0.014]	[0.013]	[0.012]
Female	0.011	-0.029***	-0.068***	-0.011
	[0.007]	[0.008]	[0.007]	[0.007]
Age	0	-0.003	0.001	-0.002
	[0.002]	[0.003]	[0.002]	[0.002]
Age Squared	-0.001	0.001	-0.003	0.001
	[0.003]	[0.003]	[0.003]	[0.002]
Mother's Education				
Less than High School	-0.003	0.002	-0.011	-0.019
	[0.012]	[0.014]	[0.013]	[0.012]
Some Post Secondary	-0.011	-0.026**	-0.011	-0.013
	[0.010]	[0.012]	[0.011]	[0.010]
BA or More	0.014	-0.002	0.006	0.005
	[0.015]	[0.016]	[0.019]	[0.014]
None Reported	0	-0.003	0.007	-0.046**
	[0.020]	[0.023]	[0.021]	[0.019]
Father's Education				
Less than High School	0.002	0.012	0.006	0.012
	[0.013]	[0.015]	[0.015]	[0.013]
Some Post Secondary	-0.012	-0.005	-0.011	0.001
	[0.010]	[0.011]	[0.012]	[0.010]
BA or More	-0.005	-0.006	-0.006	-0.003
	[0.014]	[0.016]	[0.015]	[0.014]
None Reported	0.009	0.019	-0.01	0.003
	[0.020]	[0.022]	[0.022]	[0.019]
Immigrant Mother	-0.01	-0.019	-0.013	-0.012
	[0.011]	[0.013]	[0.013]	[0.012]
Immigrant Father	0.004	0.008	0.002	0
	[0.010]	[0.012]	[0.012]	[0.010]

Constant	4.801***	4.690***	4.789***	4.848***
	[0.119]	[0.138]	[0.127]	[0.115]
Observations	12370	12370	12370	12370
R-squared	0.26	0.16	0.33	0.24

Notes:
1. Standard errors in parentheses. *, **, *** statistically significant at 10%, 5%, 1% level.

Table 3: First Stage Results for Years of Schooling

Variable	IV1	IV2
CSLA>=15 & CSLA<16	0.630***	0.494***
	[0.113]	[0.119]
CSLA>=16	0.578***	0.479***
	[0.125]	[0.133]
CSEA=5 or 6	-	0.222
		[0.179]
CSEA=7	-	0.389**
		[0.173]
Female	0.316***	0.319***
	[0.075]	[0.075]
Age	0.162***	0.158***
	[0.012]	[0.012]
Age Squared	-0.193***	-0.188***
	[0.012]	[0.013]
Mother's Education		
Less than High School	-0.740***	-0.729***
	[0.111]	[0.112]
Some Post Secondary	0.361***	0.364***
	[0.117]	[0.117]
BA or More	0.135	0.14
	[0.173]	[0.172]
None Reported	-1.313***	-1.293***
	[0.188]	[0.188]
Father's Education		
Less than High School	-0.809***	-0.818***
	[0.109]	[0.109]
Some Post Secondary	0.220*	0.207
	[0.129]	[0.129]
BA or More	0.564***	0.557***
	[0.147]	[0.147]
None Reported	-1.252***	-1.262***

	[0.170]	[0.169]
Immigrant Mother	0.446***	0.435***
	[0.123]	[0.123]
Immigrant Father	0.241**	0.231*
	[0.123]	[0.123]
Constant	9.211***	9.132***
	[0.290]	[0.314]
Observations	12370	12370
R-squared	0.27	0.27
First Stage F-Statistics	15.78	9.96

Notes:
1. Standard errors in parentheses. *, **, *** statistically significant at 10%, 5%, 1% level.

Table 4: Earnings Regressions

Variable	OLS1	OLS2	IV2	IV3	IV3
Years of Schooling	0.085***	0.070***	0.089***	0.084***	0.054***
	[0.004]	[0.005]	[0.026]	[0.005]	[0.015]
Female	-0.414***	-0.411***	-0.416***	-0.414***	-0.408***
	[0.024]	[0.023]	[0.026]	[0.024]	[0.024]
Experience	0.067***	0.066***	0.067***	0.067***	0.065***
	[0.004]	[0.004]	[0.006]	[0.004]	[0.004]
Experience Squared	-0.001***	-0.001***	-0.001***	-0.001***	-0.001***
	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]
Average Literacy Score		0.003***			0.005**
		[0.000]			[0.003]
Constant	4.657***	4.115***	4.610***	4.670***	3.548***
	[0.082]	[0.115]	[0.322]	[0.085]	[0.553]
Observations	7569	7569	7569	7569	7569
R-squared	0.39	0.41	0.39	0.39	0.39
First Stage F-Statistics					
Years of Schooling			20.37	139.73	139.73
Average Literacy Score					16.84

Notes:
1. Standard errors in parentheses. *, **, *** statistically significant at 10%, 5%, 1% level.
2. Parental characteristics, province of residence and urban/rural status dummies are included as control variables in both OLS and IV regression.

Table 5: Quantile Earnings Regressions

	10th Quantile	25th Quantile	Median	75th Quantile	90th Quantile
Female	-0.48***	-0.46***	-0.36***	-0.36***	-0.36***
	(0.048)	(0.032)	(0.020)	(0.021)	(0.032)
Years of Schooling	0.079***	0.065***	0.069***	0.067***	0.057***
	(0.011)	(0.0065)	(0.0041)	(0.0042)	(0.0069)
Experience	0.088***	0.084***	0.059***	0.050***	0.042***
	(0.0063)	(0.0041)	(0.0025)	(0.0024)	(0.0036)
Experience Squared	-0.0015***	-0.0015***	-0.001***	-0.0008***	-0.0006***
	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)
Average Literacy Score	0.0022***	0.0032***	0.0029***	0.003***	0.0028***
	(0.0007)	(0.0004)	(0.0003)	(0.0003)	(0.0005)
Constant	3.33***	3.67***	4.26***	4.62***	5.12***
	(0.19)	(0.12)	(0.083)	(0.079)	(0.13)
Observations	7768	7768	7768	7768	7768

Standard errors in parentheses. *, **, *** statistically significantly different from zero at the 10%, 5% and 1% level of significance. All regressions include controls for province of residence and urban/rural status