June 1, 2000

Catalogue no. 89F0120XIE

Schooling, Literacy and Individual Earnings

by Lars Osberg

Abstract

How much of the economic benefit of an education can be attributed to literacy skills?

Using Canadian data from the International Adult Literacy Survey (IALS), this paper examines the validity of averaging literacy scores across jurisdictions. It also sets bounds on the proportion of the individual benefit of education that can be explained by level of literacy. Osberg focuses on the methodology behind measures of skills such as literacy and, in so doing, opens an important methodological debate.

Direct measures of skill attainment, like the IALS, are used to assess the importance of educational outcome skills—in this case, literacy—in determining labour market outcomes, such as earnings. Policy makers also use them to direct resources most efficiently. These measures of skill are the product of complex statistical procedures. This paper compares the strength of IALS measures with other approaches to estimating the impact of literacy on individual earnings.

Osberg shows that for men employed full time and full year, for example, literacy accounts for about 30% of the economic return from education. Whatever way the literacy score is stretched for the full-time, full-year work force, it is always statistically significant. Yet the contribution of literacy can explain 40% to 45% of this return when the question is turned around to ask: what is the maximum fraction of the economic benefit of an education that can be explained by the inclusion of measured literacy skills?

Introduction

When assessing the average quality of public services, we can often determine whether one outcome is preferable to another, but we cannot usually give a unique ratio to that preference. Measures of outcome quality are often both crucial to program evaluation and highly problematic, since such numbers are typically ordinal numbers. In other words, they allow us to assign a rank: A is better than B. They are not, however, cardinal numbers, which would allow us to assign rank and magnitude: to say, for example, that ‘2’ is twice as far away from ‘0’ on the number line as is ‘1.’

Measures of educational outcomes—such as school grades or literacy scores—provide a particularly important example. Although we may all agree that an A is better than a B, which in turn dominates a C, we may well disagree on how to average an A and a C. Although tests of literacy skills can be used to assess whether one individual is more literate than another, the statement that one individual is *n*% more literate than another makes little sense. Nevertheless, it has become commonplace to make comparisons between the average literacy levels of different jurisdictions. These comparisons are often used to justify appeals for increased public spending, on the basis that a more literate work force will be more productive. Thus, in order to make a cost–benefit calculation, we must ask, “How much does literacy matter?”—or, more precisely, as this paper asks, “How much of the economic benefit of an education can be attributed to literacy skills?”

The literature on the economic benefits of education is vast, but most of it has attempted to measure the economic benefits of investment into the educational process. The economic literature that assesses the benefits of education has not generally used direct measures of outcomes—such as literacy—to explain earnings. It might be very useful to know which of the skills produced by the educational system actually pay off, and how large that payoff is. There is also good reason to be cautious when using measures of student skill attainment as explanatory variables to predict earnings. Input measures such as years of schooling or the amount of money spent per pupil have natural metrics (for example, years or dollars) that are cardinal numbers and therefore easily averaged. Scores from the IALS—like many other social and cognitive skill scores—are ordinal numbers that tell us one is higher than another, but not by how much.

This paper uses Canadian data from the International Adult Literacy Survey to examine the validity of averaging literacy scores across jurisdictions, and to set bounds on the proportion of the individual benefit of education that can be explained by level of literacy.

In focusing on the methodology behind measures of skills such as literacy, this report opens an important methodological debate. We acknowledge that its use of specialized, mathematical language to construct its argument may limit the appeal of the paper to the wider literacy audience, and that the views presented in it do not necessarily reflect those of the granting agencies or reviewers.

Using ordinal test scores in mathematical models

The mathematical issue in this paper is the use of ordinal literacy scores as one of several potential explanatory variables. We want to know the relative *size* of its impact in explaining a cardinal variable (earnings), while we also take into account the influence of other factors, such as work experience. The standard reflex of most empirical economists would be to use the literacy score of an individual to help explain individual earnings. Unfortunately, the use of any particular scaling of ordinal scores is open to criticism since ordinal scores do not have a natural metric. In other words, the scores do not have a consistent and regular scale.

Our knowledge of literacy may allow us to exclude some possible manipulations of literacy scales as unreasonable. For example, we may not be sure how a score at the top end of the literacy distribution compares with a score at the mid-point. If the median individual scores 250, and the 95th-percentile individual scores 400, it may not be accurate to say that the latter individual has 60% more literacy than the former. We could scale the same test to produce scores of 100 and 900, which might also be thought reasonable. However, a rescaling which produced the values of 100 and 100,000,000,000 would be thought by most observers to be unreasonable.

If we can limit the set of reasonable scaling manipulations, we may be able to create an accurate empirical measure of the impact of literacy on earnings. However, assessing the range of reasonable transformations of IALS scores requires some discussion of how those scores are generated.

The International Adult Literacy Survey

Rather than labelling individuals ‘literate’ or ‘illiterate,’ modern work on literacy describes a continuum of literacy skills in the population. The IALS therefore aims at assessing the degree of functional literacy—that is, the ability of individuals to use written information in real life situations. Individuals were presented with realistic situations of graduated difficulty to assess their competency. Some of the test tasks include the following:

- balancing a chequebook, calculating a tip, and completing an order form (quantitative literacy);
- locating and using information from documents such as job applications, payroll forms, transportation schedules, maps and tables (document literacy); and
- understanding and using information from texts such as editorials, news stories, poems and fiction (prose literacy).

In the IALS, an individual’s literacy skills were scored on a scale of 0 to 500. While the literacy scales make it possible to compare the prose, document and quantitative skills of different populations and to study the relationships between literacy skills and various factors, the scale by themselves carry little or no meaning. That is to say, a score of 300 is not twice as high as a score of 150. The 300 score is higher, but we cannot say how much higher in terms of functional literacy.

If a questionnaire testing functional literacy is to be completed within a reasonable total time, only a limited number of questions are possible. In the IALS, there was a battery of about 33 potential test items. These questions were divided into seven blocks. Each respondent answered three blocks, that is, no more than 15 questionnaire items on each major dimension of literacy.

To assign scores to this kind of testing, statistical procedures based on Item Response Theory were employed. The potential problem with this approach is that some IALS scores may be outside the range of difficulty of the questionnaire items actually used in the survey. For example, although the easiest question in one literacy segment had a difficulty rating of 225 and the hardest was rated at 408, fully 26% of Canadian respondents scored less than 225 for the segment and 0.5% scored more than 408.

This paper lays some stress on the issue of imputed values greater than or less than the difficulty level of any question actually asked, because these scores are at the tails of the distribution of literacy proficiency. For many policy issues, it is the tails of the literacy distribution that matter. Advocates for remedial literacy programs point to the way that people with low levels of literacy are excluded from employment and normal written communication and social discourse. Their emphasis on exclusion suggests that very low literacy attainment can have qualitatively different, and very significant, employment and social impacts, compared with mid-range differences (for example, in reading speed).

On the other hand, advocates for greater educational streaming and elite programs for gifted children stress the importance of scientific and literacy excellence and of breakthrough discoveries—as exemplified by the number of Nobel Prize winners or high-tech billionaires. Such high achievers are inherently rare, however, and it is difficult to design questionnaire items to test accurately very high or very low levels of skill attainment. As a consequence, literacy tests of the general population are best able to assess variation in the middle range of literacy skills, which may not be the relevant range for many public policy debates.

Item Response Theory

The idea behind Item Response Theory is that we can see test scores probabilistically. We can draw an analogy to the probability with which high jumpers will clear a bar set at a given level: athletes of a given competency will fail occasionally at lower heights and will succeed occasionally at greater heights. If we set a specific probability of success at 80%, for example, the literacy level of an individual can be defined as the point at which individuals with that proficiency score have an 80% probability of responding correctly. This means that individuals estimated to have that score would consistently perform tasks—with an 80% probability—like others measured at that point on the scale. It also means that they will have a greater than 80% chance of performing tasks that are rated lower than their estimated proficiency on the scale.

In estimating the probability with which an individual could complete a task of given difficulty, the individual does not have to be observed performing that task. Indeed, some individuals tested in the IALS responded to as few as five questionnaire items. Proficiency scores were imputed using observed responses and background variables—for example, sex, ethnicity, language of interview, respondent education, parental education, occupation and reading practices.

Rankings of average literacy

The objective of Item Response Theory is to provide estimates of population means, and when we compare the literacy outcomes observed in different jurisdictions, the most commonly used statistic is the rank order of average scores. However, the rank order of average outcomes may be dependent on the scaling method of the scores. If alternative statistical scaling of individual literacy scores almost always results in the same rank order of average provincial literacy scores, then we might conclude that the ordinal nature of literacy scores simply does not matter much in practice. If jurisdictions often change rankings when scaling changes are made, we may wish to look deeper.

This study examines the frequency with which Canadian provinces occupy specific rankings of literacy proficiency using various mathematical transformations. For example, one way of testing for the importance of very high levels of skill is to magnify the impact of differentials in the top end of the literacy distribution by raising individual literacy scores to successively higher powers. Therefore, the first examination calculates average literacy using the five scales generated for IALS data, raised to the powers 1 through 9.

The objective is simply to assess whether alternative scales make any practical difference to provincial rankings. A total of 45 separate rankings can result from this procedure. Although a clear general tendency can be observed, there is also a significant number of reversals. For example, British Columbia is most often ranked fourth, but occasionally first or fifth.

The 45 alternative scalings of total literacy clearly show that the Canadian average literacy gradient in the total population increases generally from east to west. However, it is also clear that relative rankings within the East and the West are quite sensitive to the scaling of individual scores.

Average literacy rankings for the entire population, however, are very much a lagging indicator. People who left school 40 or even 50 years ago are mingled with recent school leavers, so it takes decades for the impact of educational policy changes to show up in the overall average. A look at the rankings of average literacy levels of the under-30 population shows little indication of an east to west gradient.

Generally then, great caution should be used in interpreting the policy implications of average literacy attainment. Based on comparison of the simple population averages of one scale, there might be a temptation to say that the education system of Atlantic Canada has 'failed' because that calculation shows average literacy levels to be lowest in Atlantic Canada. It might also be tempting to say that low literacy levels are partly responsible for the lower incomes of that region. When we look at the population under 30, however, Nova Scotia is frequently ranked as having the *highest* average literacy level in the country. It seems, then, that caution is in order. Before policy conclusions are drawn, we need to be sure that rankings of average proficiency are not just artifacts of the scaling of individual scores.

Literacy and earnings

Literacy is one of the major objectives of the educational system, and the number of years of education has long been found to be a good predictor of individual earnings. How much of the benefits of education can be accounted for by an individual's level of literacy?

To assess this issue, we must choose a plausible measure of literacy. IALS data present five plausible values for literacy scores. However, no actual test items were used with very low or very high difficulty levels. Since scores that were assigned above or below the skill levels actually tested are solely the result of statistical imputation, we might wonder whether such scores should be truncated, and how.

All these issues—truncation, power transformations and rank information—are measurement choices, but choices are not limited to the appropriate measure of literacy. For example, some would argue that estimation of the effects literacy and education have on earnings should be restricted to full-time, full-year workers (to have some control for hours of labour supply). But, if education or literacy enables individuals to gain access to employment, it could also be argued that the sample should include all workers.

Does measurement choice matter?

The strategy of this paper is to present the very simplest human capital earnings equation and to

1. add alternative possible measures of literacy; and
2. experiment with alternative subpopulations and measurements of education.

The first examination of men employed full time and full year shows that literacy accounts for about 30% of the return to education. Whatever way the literacy score is stretched for the full-time, full-year work force, it is always statistically significant. Only by strongly accentuating the relative importance of differentials at the top end of the literacy scale (by raising the literacy score to the nth power) can we reduce the estimated impact of literacy on the return to schooling to about one-sixth.

If we look at males who work full time, full year, and measure education by credentials obtained, the conclusion that literacy skills explain a significant fraction of the return to education is altered. In some cases, the impact of literacy skills appears greater. It appears that including a control for measured literacy skills reduces by 40% to 45% the estimated benefit of a university education. Although the impact of including measured literacy with very low education is less (a 16% to 26% decline), this examination still indicates that much of the measured benefit of education is due to literacy skills.

The last examination provides a cautionary note. For female full-time and part-time workers, literacy scores are never statistically significant. Indeed, it is possible to find a rescaling of individual literacy scores such that the estimated return to education, net of literacy, rises when literacy scores are included as an explanatory variable. This is not a very plausible result.

Over the range of measurement choices considered, literacy skills generally explain a significant part of the return to education—but not always. This suggests that we could turn the question around and ask, “What is the maximum fraction of the return to education that can be explained by the inclusion of measured literacy skills?” The range of estimated returns to years of schooling when various measurement choices about literacy scoring are made, compared with the baseline estimated of the return to years of schooling when literacy scores are not considered, suggests that increased literacy is only part of the reason why education pays. The estimates show extreme variation, highlighting the uncertainty about how much of the financial return to years of education is explained by literacy skills. Whatever the rescaling of literacy scores, however, it is very hard to push the contribution of literacy above 40% to 45% of the return to education.

Discussion

When a large number of possible combinations of measurement choices exist, meta-analysis can detect patterns in the implications of measurement choices. Using regression models, several combinations of factors were examined to explain the estimated rate of return to years of education. The results using a different scale of plausible scores (five scales were developed for IALS) increased the estimated rate of return per year of education by 0.7%.

By raising the literacy score to successively higher powers, individual literacy scores increasingly emphasize the importance of differences at the top end of the literacy distribution. For full-time, full-year males, magnifying the importance of literacy differentials at the top end of the literacy distribution has a statistically significant effect on the measured return to education. In the meta-analysis, entering the power to which literacy scores are raised and the square of that power can be interpreted to mean that the estimated rate of return to education is minimized. Conversely, a logarithmic transform of literacy scores compressing the influence of differentials at the top end of the literacy distribution appears to inflate the measured influence of years of schooling on earnings. So, in addition to the fact that the specific plausible scoring measurements used typically matter for the estimated rate of return to education (conditional upon literacy), it also matters whether the scaling of literacy emphasizes differences at the top end of the distribution, relative to the bottom.

It also seems to matter how we treat very low literacy scores. In one examination, all literacy scores below the minimum level of difficulty actually asked are raised to the minimum score level tested. A second examination, on the other hand, sets to zero the literacy score of any person assigned a score less than the least difficult question item. Implicitly, it accentuates measured literacy differences in the bottom tail, and such a choice measurably improves the estimated return to education.

Conclusion

The development of direct measures of skill attainment, such as the IALS data, offers labour economists a powerful new tool to help explain labour market outcomes. There is a great deal of useful information in such test scores. This paper demonstrates that literacy test scores have a statistically and empirically significant relationship with individual earnings, and that the effect is valid for a large variety of measurement choices.

Nevertheless, some caution is in order when using direct measures of skill attainment in statistical analysis. Over the years, labour economists have developed many complex statistical techniques for working with data but, in general, the underlying concepts have been clearly observable magnitudes—such as number of children, hourly wage or marital status—that can be measured either as cardinal numbers, or as discrete states.

However, literacy is a complex concept for which there is no natural unit of measurement. Although direct measures of literacy proficiency like IALS can rank individuals in literacy attainment, literacy scores are the products of complex statistical procedures. These procedures involve many of the same variables—such as education or age—that labour economists would usually expect to play an independent role in determining labour market outcomes. Literacy scores are also inherently ordinal numbers, and a variety of transformations of those scores may be equally plausible. The method of calculation of literacy scores may therefore be very important for the perceived impact of literacy on labour market outcomes. More generally, since many public services have a quality dimension that is similarly complex and inherently ordinal, the method of calculation and the scaling of quality measures of public sector outcomes may become central to policy debates.

These highlights summarize the seventh International Adult Literacy Survey monograph, **Schooling, Literacy and Individual Earnings,** Statistics Canada Catalogue number 89-552-MPE, no. 7. For further information, contact T. Scott Murray at (613) 951-9035.