

# Testing the Internal Validity of Compulsory School Reforms as Instrument for Years of Schooling\*

by

Giorgio Brunello

(University of Padua)

Margherita Fort

(University of Bologna)

Guglielmo Weber

(University of Padua)

Christoph T. Weiss

(European Investment Bank)

## ABSTRACT

In the large empirical literature that investigates the causal effects of education on outcomes such as health, wages and crime, it is customary to measure education with years of schooling, and to identify these effects using the exogenous variation provided by school reforms increasing compulsory education and minimum school leaving age. If these reforms are correlated to changes in school quality, and school quality is an omitted variable, this identification strategy may fail. We test whether this is the case by using the information provided by two distinct test scores on mathematics and reading and find that we cannot reject the internal validity of this popular identification strategy.

**JEL:** C26, I2

**KEYWORDS:** human capital, instrumental variables, nested models

Corresponding author: Margherita Fort, Department of Economics, University of Bologna, Piazza Scaravilli, Bologna; e-mail: [margherita.fort@unibo.it](mailto:margherita.fort@unibo.it)

---

\* We are grateful to Daniele Fabbri, Sergio Pastorello, Hashem Pesaran, Lorenzo Rocco, Richard Spady, Uwe Sunde, Frank Windmeijer and to the audiences in Gothenburg (EEA) and Turin (EALE) for comments and suggestions. This paper was written while Christoph T. Weiss was Antonveneta Fellow at the Department of Economics of the University of Padua. Financial support from CSEA, the Antonveneta Centre for Economic Research, is gratefully acknowledged. Giorgio Brunello is affiliated with LUMS and IZA; Margherita Fort is affiliated with IZA. The views expressed in this paper are those of the authors and do not necessarily represent those of the EIB or EIB policy. All errors are our own.

## Introduction

In the large literature on the returns to education, the vast majority of empirical studies uses the number of years of schooling to measure individual education. The reliance on years of schooling is understandable because this indicator is relatively easy to compute and widely accessible, but this advantage comes at the price of neglecting additional and perhaps equally important dimensions of education, including the quality of learning. In their analysis of the effects of education on economic growth, Hanushek and Woessmann (2012) point out that using years of schooling in cross-country studies is equivalent to assuming that an additional year of schooling delivers the same increase in knowledge and skills regardless of the education system.<sup>1</sup>

In this literature, it has become customary since Acemoglu and Angrist (2001) to identify the causal effects of education on various outcomes – including crime (Lochner and Moretti, 2004, and Machin et al., 2011), earnings (Oreopoulos, 2006), fertility (McCrary and Royer, 2011 and Fort et al., 2011), health (Lochner, 2011), intergenerational transmission of human capital (Black et al., 2005), migration (Machin et al., 2012), mortality (Lleras-Muney, 2005, and Clark and Royer, 2013), obesity (Brunello et al., 2013) and voting behaviour (Milligan et al., 2004) – with the assumption that changes in minimum school leaving age or in years of compulsory education affect outcomes only by affecting years of schooling.

There are reasons to suspect, however, that minimum school leaving age reforms might have affected not only the quantity but also the quality of education, which is typically omitted from regressions relating outcomes to years of schooling.<sup>2</sup> First, by increasing enrolment and forcing less talented individuals to remain in school longer, these reforms may have changed the ability pool of students, with consequences for education and other outcomes.<sup>3</sup> Second, exogenous changes in compulsory education have often been accompanied, at least in Europe, with qualitative changes that have altered the length and characteristics of school tracking.<sup>4</sup> Last but not least, reforms may have affected at least temporarily pupil-teacher ratios,

---

<sup>1</sup> Another implicit assumption often made in this literature is that formal schooling is the primary source of education and that variations in the quality of non-school factors exert a negligible effect on education outcomes. As pointed out by Feinstein et al. (2006), this assumption needs not hold and lifelong learning effects beyond the stage of higher education during emerging adulthood may be substantial.

<sup>2</sup> The internal validity of compulsory school reforms as an instrumental variable for schooling is questioned by Holmlund, Lindhal and Plug (2011). In their Handbook chapter, Black and Devereux (2011, p.1526) write that “...changes in compulsory education laws tend to accompany curriculum changes, new buildings, and the hiring of new teachers. Any of these could have direct effects. Also, college openings may respond to educational demand in the local area. As a result, despite the often significant efforts of the researchers, the credibility of the estimates can always be called into question...”

<sup>3</sup> See Sacerdote (2011) for a review of the large literature on peer effects in education.

<sup>4</sup> Meghir and Palme (2005) for instance, examine the effects of a compulsory school reform in Sweden in the 1950s, which increased minimum school leaving age by two years and delayed tracking at the same time. Bhuller et al. (2011) describe the Norwegian school reform which extended years of compulsory education from 7 to 9 years, and unified at the same time the education system beyond age 15/16. The Italian school reform of 1962 added three years of compulsory education (from 5 to 8)

because building new schools and hiring new teachers to accommodate the increased population of students induced to stay in school longer typically takes time.

If quality affects outcomes – as suggested by Hanushek and Kimko (2000) and Barro and Lee (2001) – but is omitted from regressions relating these outcomes to years of schooling because of lack of data, the possibility that compulsory school reforms have an impact not only on years of schooling but also on school quality can invalidate the exclusion restriction at the root of the instrumental variable (IV) strategy adopted in the recent literature, which imposes that compulsory school reforms affect outcomes only by changing years of schooling.

In this paper, we propose a simple test of the validity of this exclusion restriction, which requires two *distinct* measures of cognitive test scores and an instrumental variable (years of compulsory schooling). Alternative strategies that use more than one instrument for identification can of course be designed. However, since it is quite hard to find credible instruments for both years of schooling and school quality, the requirement of a single instrument makes our test – we believe – particularly appealing.

Conditional on the additional assumption that the quantity of education affects both test scores in the same way – which is met in our data – we show that the ratio of the IV estimates of the effect of each test score on the outcome of interest is equal to one under the null hypothesis that the exclusion restriction is valid, (i.e. compulsory schooling reforms do not affect school quality) and different from one under the alternative hypothesis that the exclusion restriction fails to hold.<sup>5</sup> Very conveniently, the test can be implemented with an auxiliary regression that does not require data on the outcome variable (earnings, health, crime or else).

We use individual data from seven European countries (Denmark, Finland, Great Britain, Ireland, Italy, the Netherlands and Northern Ireland) on prose and quantitative literacy test scores provided by the International Adult Literacy Survey (IALS) and country-specific information on compulsory school reforms to implement our test and find that the null hypothesis is not rejected by our data. Using simulation analysis, we show that our test has power to reject the null even when the covariance between the instrument and school quality is small. We also show that our results are not sensitive to small deviations from the assumption that the quantity of education has the same effect on both prose and

---

and eliminated at the same time the distinction between vocational and academic tracks in lower secondary education. By postponing the time when students can choose between vocational and academic education, these reforms may have influenced average school quality insofar as different tracks provide different type and quality of education (see the discussion in Brunello and Checchi, 2007).

<sup>5</sup> The test does not rely on the empirical counterpart of the orthogonality conditions (Sargan, 1958) nor rests on the comparison of the behaviour of different estimators for the same parameter under the null or the alternative hypothesis (Durbin, 1954, Wu, 1973 and Hausman, 1978) nor does it require binary instruments and binary treatments (Kitagawa (2008), Huber (2013) and Huber and Mellace (2013)).

quantitative literacy scores. We conclude that our results support the current widespread practice of using compulsory school reforms to estimate the causal effects of years of schooling on several outcomes.

The paper is organized as follows. Section 1 sets up the problem and Section 2 describes the test. In Section 3, we introduce the data and in Section 4 we present our results. Using simulations, we investigate the power of the test in Section 5. Conclusions and an Appendix follow.

## 1. Setting Up the Problem

Following Hanushek and Woessmann (2012), we define human capital as the stock of cognitive skills. Let outcome  $Y$  (e.g. wages, health or crime) be a function of human capital  $H$ <sup>6</sup>

$$Y = \pi H + \omega \tag{1}$$

where  $\pi \neq 0$ , additional covariates and the constant term have been partialled out and  $\omega$  is a disturbance term which includes unobserved ability. Since the covariance between  $H$  and  $\omega$  is different from zero (see Card, 1999), the identification of the causal effect of  $H$  on  $Y$  requires a source of exogenous variation, which affects  $H$  but is orthogonal to  $\omega$ . Let this source be  $Z$ , with  $Cov(Z, \omega) = 0$ , and let human capital  $H$  be a linear function of years of schooling  $S$ , school quality  $Q$  and a residual  $\varepsilon$ <sup>7</sup>

$$H = \lambda S + \theta Q + \varepsilon \tag{2}$$

where  $\lambda > 0$ ,  $\theta > 0$  and the error term  $\varepsilon$  includes unobserved factors such as ability and labour market experience.

In most of the empirical literature, school quality is omitted or unobserved and human capital is measured with years of schooling  $S$ . Substituting equation (2) into (1) yields

$$Y = aS + \mu \tag{3}$$

---

<sup>6</sup> Hanushek and Woessmann (2012) find that cognitive skills – measured by test scores – significantly affect growth. Heckman, Stixrud and Urzua (2006) find that schooling, employment, work experience and choice of occupation are affected by the same latent non-cognitive and cognitive skills. Green and Riddell (2003) use IALS data for Canada and find that cognitive skills contribute significantly to earnings and that their inclusion in earnings equations reduces the measured impact of schooling. Murnane, Willett and Levy (1995) use data from two longitudinal surveys of American high school seniors and show that basic cognitive skills had a larger impact on wages for 24-year-old men and women in 1986 than in 1978.

<sup>7</sup> In Hanushek and Woessmann (2012), the residual includes family inputs, health and labour market experience. In their setup, the quantity and quality of education affect human capital in a multiplicative way. In the Appendix, we extend equation (2) by adding the product between  $S$  and  $Q$  and show that our testing strategy is unaffected.

where  $a = \pi\lambda$  and  $\mu = \pi\theta Q + \pi\varepsilon + \omega$ . The causal effect of  $S$  on outcome  $Y$  is estimated using the exogenous variation in compulsory education as instrumental variable for schooling. This strategy produces consistent estimates of the returns to schooling if  $Cov(Z, \mu) = \pi\theta Cov(Z, Q) + \pi Cov(Z, \varepsilon) = 0$ .

## 2. The Test

We design a test to verify whether the exclusion restriction that compulsory school reforms affect outcomes only indirectly by influencing years of schooling is valid. A key requirement for this test is the availability of two distinct measures of human capital  $H$  other than schooling  $S$ ,  $T_1$  and  $T_2$ , where  $T$  is a standardized test score. Let  $T_1$  be a measure of reading achievement and  $T_2$  a measure of mathematics achievement. In the case of adults, these scores reflect both the quantity and quality of their schooling as well as their learning on the job and in the labour market. A measure of latent human capital  $H$  more encompassing than years of schooling  $S$  is the average test score  $\frac{T_1 + T_2}{2}$ , which combines the information on reading and math skills and depends both on years of schooling and on school quality.<sup>8</sup> Using this score as the measure of  $H$ , we write<sup>9</sup>

$$Y = \psi \frac{T_1 + T_2}{2} + \xi \quad [4]$$

where  $\psi > 0$ . Let the relationship between test scores, years of schooling and school quality be linear and given by

$$T_1 = \lambda_1 S + \theta_1 Q + \varepsilon_1 \quad [5a]$$

$$T_2 = \lambda_2 S + \theta_2 Q + \varepsilon_2 \quad [5b]$$

where  $\varepsilon_1$  and  $\varepsilon_2$  are residual terms. Solving equations [5a] and [5b] for  $S$  yields<sup>10</sup>

---

<sup>8</sup> According to Hanushek and Woessmann (2012, pp.271-272), “human capital is [...] a latent variable that is not directly observed. To be useful and verifiable, it is necessary to specify the measurement of  $H$ . The vast majority of existing theoretical and empirical work on growth begins – frequently without discussion – by taking the quantity of schooling of workers ( $S$ ) as a direct measure of  $H$ . A more compelling alternative is to focus directly on the cognitive skills component of human capital and to measure  $H$  with test-score measures of mathematics, science, and reading achievement”.

<sup>9</sup> The disturbance term  $\xi$  includes both unobserved ability and (classical) measurement error.

<sup>10</sup> Adding a multiplicative term in  $S$  and  $Q$  does not alter the test, as long as we are prepared to assume that  $Q$ ,  $S$  and  $Z$  follow a multivariate normal distribution. See the Appendix for details.

$$T_2 = \delta T_1 + \varrho_1 Q + \tilde{\varepsilon} \quad [6]$$

where  $\delta = \lambda_2 / \lambda_1 > 0$ ;  $\varrho_1 = \theta_2 - \delta\theta_1$ ;  $\tilde{\varepsilon} = \varepsilon_2 - \delta\varepsilon_1$ .

The two test scores provide distinct information on human capital if  $\delta \neq 1$  and/or  $\varrho_1 \neq 0$ . Using equation (6) to replace  $T_2$  in equation (4) we obtain

$$Y = \beta T_1 + v_1 \quad [7]$$

where  $\beta = \frac{\psi}{2}(1 + \delta)$ ;  $v_1 = \frac{\psi}{2}\rho_1 Q + \frac{\psi}{2}\tilde{\varepsilon} + \xi$ . In a similar fashion, we can use equation (6) to replace  $T_1$  in equation (4) and get

$$Y = \gamma T_2 + v_2 \quad [8]$$

where  $\gamma = \frac{\psi}{2}\left(1 + \frac{1}{\delta}\right)$ ;  $v_2 = -\frac{\psi}{2}\frac{\rho_1}{\delta}Q - \frac{\psi}{2}\frac{1}{\delta}\tilde{\varepsilon} + \xi$ . Finally, placing equations (5) and (6) in equation (4) we get equation (3), where  $\alpha = \frac{\psi}{2}(1 + \delta)\lambda$  and  $\mu = \frac{\psi}{2}[(1 + \delta)\theta_1 + \rho_1]Q + \frac{\psi}{2}(1 + \delta)\varepsilon_1 + \frac{\psi}{2}\tilde{\varepsilon} + \xi$ .

We make the following assumption, which we maintain throughout:

*Assumption 1:*  $Cov(Z, \varepsilon_1) = 0$ ;  $Cov(Z, \tilde{\varepsilon}) = 0$ ;  $Cov(Z, \xi) = 0$ .

This assumption states that compulsory education  $Z$  is uncorrelated with both unobserved ability and measurement error in equation (4) and affects test scores only insofar as it changes years of schooling  $S$  and school quality  $Q$ .

We want to test the internal validity of the instrumental variable  $Z$  in regression (3). Under Assumption 1, internal validity holds when the following condition is met (null hypothesis):

$$H_0 : \frac{\psi}{2}[(1 + \delta)\theta_1 + \rho_1]Cov(Z, Q) = 0 \quad [9]$$

Notice that  $Cov(Z, Q) = 0$  is both a necessary condition for  $H_0$  and a sufficient condition for the instrument to be valid for  $S$  in equation (3) under Assumption 1.

To derive the test statistic, we consider the ratio between the IV estimators of parameters  $\beta$  and  $\gamma$  in equations (7) and (8). When the null hypothesis does not hold, and under Assumption 1, this ratio is given by

$$\frac{\beta_{IV}}{\gamma_{IV}} = \frac{\frac{Cov(Y, Z)}{Cov(T_1, Z)}}{\frac{Cov(Y, Z)}{Cov(T_2, Z)}} = \frac{Cov(T_2, Z)}{Cov(T_1, Z)} = \frac{\delta Cov(T_1, Z) + \rho_1 Cov(Z, Q)}{Cov(T_1, Z)} \quad [10]$$

Given Assumption 1, under the null hypothesis  $Cov(Z, Q) = 0$ , the ratio becomes

$$\frac{\beta_{IV}}{\gamma_{IV}} = \frac{Cov(T_2, Z)}{Cov(T_1, Z)} = \delta \quad [11]$$

A particularly convenient condition to make the proposed test statistic fully operational is the following:

*Assumption 2* (Distinct effects of school quality on test scores):  $\delta = 1$ ;  $\rho_1 \neq 0$ .

In words, Assumption 2 says that when years of schooling  $S$  have the same effect on test scores ( $\delta = 1$ ), the ratio of covariances in equation (11) is equal to 1 under the null hypothesis. In this particular case, the two test scores are distinct only because of the different effect of school quality ( $\rho_1 \neq 0$ ). Notice that the ratio between  $Cov(T_2, Z)$  and  $Cov(T_1, Z)$  is the IV estimate of the effect of  $T_1$  on  $T_2$ . Therefore, our test can be implemented simply by regressing  $T_2$  on  $T_1$ , using  $Z$  as the instrumental variable for  $T_1$ . Particularly attractive features of this test are that it does not require information on the outcome variable  $Y$ , and that only one instrument is required for identification.<sup>11</sup>

In the sequel, we shall estimate equations (5a) and (5b) and show that Assumption 2 holds in at least one well-known and widely used data set, the International Adult Literacy Survey, which contains information on  $S$ ,  $T_1$ ,  $T_2$  and can be enriched with a proxy for  $Q$ . And we shall also show, by means of simulations, that small departures from Assumption 2 do not affect the power of the test.

---

<sup>11</sup> If we had two credible instruments, we could regress the outcome  $Y$  on years of schooling  $S$  and school quality  $Q$ , instrumenting both covariates. This estimator would be consistent both under the null ( $Cov(Z, Q) = 0$ ) and under the alternative hypothesis ( $Cov(Z, Q) \neq 0$ ).

But the question arises of whether Assumption 2 can be verified in data sets with no information on school quality, that are commonly used in applied work on the returns of education. After all, our test statistic is meant to help investigators decide whether they really need to collect information on  $Q$ , or they can instead neglect the effects of changes in  $Q$  when estimating the causal effect of schooling,  $S$ , on a chosen outcome variable,  $Y$ . The answer to this question is a qualified yes. If we are prepared to assume that school quality is constant within pre-defined groups (such as schools, or classes, or school districts, at a given point in time or over a period of time), we can estimate directly an equation like (6) by exploiting only within cell variability. This of course identifies  $\delta$ , but is uninformative on  $\varrho_i$ . But given that we can produce a consistent estimate for  $\delta$  irrespective of the value taken by  $\varrho_i$ , a simple test for the condition  $\varrho_i \neq 0$  is readily available. In fact, if  $\varrho_i = 0$  holds,  $Q$  does not enter equation (6), and a standard GLS “random effects” estimator for  $\delta$  is both consistent and efficient. If  $\varrho_i \neq 0$ , instead, GLS is inconsistent (while the within group estimator retains its consistency). Thus a simple Hausman test of random effects versus fixed effects can be used to test for  $\varrho_i \neq 0$ .<sup>12</sup>

### 3. The Data

The data requirements for the implementation of the test include: a) two distinct test scores; b) the availability of these scores for several cohorts of individuals, some affected and some not affected by compulsory school reforms. These requirements exclude well known datasets, such as the Programme for International Student Assessment (PISA), the Progress in International Reading Literacy Study (PIRLS) and the Trends in International Mathematics and Science Study (TIMSS) because the standardized tests in these data refer only to one or two cohorts of students. A dataset that meets our requirements is the International Adult Literacy Survey (IALS onwards), which was carried out by Statistics Canada and the OECD in a sub-sample of OECD countries and in three different waves – 1994, 1996, 1998 – using a common questionnaire, with the purpose of collecting comparable information on adult literacy.

Literacy in IALS has three dimensions: prose, document and quantitative. Prose literacy is defined as the knowledge and skills needed to understand and use information from texts including editorials, news stories, poems and fictions. Document literacy pertains to the knowledge and skills required to locate and use information contained in various formats, including job applications, payroll notices, transportation schedules, maps, tables and graphics. Quantitative literacy is defined as the knowledge required to apply arithmetic operations, either alone or sequentially, to numbers embedded in printed materials, calculating a

---

<sup>12</sup> Here we implicitly assume consistency of the within group estimator. This is equivalent to assuming that the error term in (6), that is the quasi-difference between unobserved determinants of  $T_1$  and  $T_2$  in (5a) and (5b), is uncorrelated with  $T_i$ . Alternatively, an instrument displaying within group variability is required – but this affects the type of cells that one can consider.



tip, completing an order form, or determining the amount of interest on a loan.<sup>13</sup> Results of the tests are scaled in the range (0,500).

We focus on prose and quantitative literacy tests, which are close enough to reading and mathematics tests. Cognitive skills as measured by IALS test scores reflect both the formal education process – its quantity and quality – and the learning activities taking place after education is completed. They are therefore a good proxy of the stock of cognitive human capital accumulated by each individual until the time of the interview. For each selected country, we retain males and females between age 25 and 60 at the time of the survey.

IALS was carried out in 18 countries, from Europe, the Americas and Oceania. We exclude from our sample the US and Germany, because minimum school leaving age in these countries took place at the state or regional level and we do not have information in our data on the region/state of birth or residence. We also exclude other non-European countries and focus only on the relatively homogeneous group of Western European countries, which includes Denmark, Finland, Great Britain, Ireland, Italy, the Netherlands, Northern Ireland, Norway and Sweden.<sup>14</sup> Finally, we exclude Switzerland and Belgium either because of the lack of information on compulsory school reforms or because reforms took place too late.<sup>15</sup>

We pool data from the selected group of European countries to exploit the cross-country variation in minimum school leaving age and to distinguish this variation from a pure cohort effect.<sup>16</sup> Pooling together data from several countries may raise concerns if the key effects are heterogeneous across countries. To verify whether there is evidence of heterogeneity, we regress quantitative (prose) literacy on prose (quantitative) literacy – using compulsory school reforms as instrument for the latter – and allow the coefficient of prose (quantitative) literacy to vary by country. We then select the sub-sample of countries for which we could not reject the hypothesis that the effect of prose (quantitative) on quantitative (prose) literacy is the same across countries. This leads us to exclude from our final sample both Norway and Sweden.

The raw data show both the positive association between education and test scores and the substantial variation in test scores for a given number of years of schooling (see Figure 1), which can be due to country and cohort effects, and to differences in parental background, school quality and learning from labour market experience. Figure 2 plots quantitative versus prose literacy scores and shows that: a) the two

---

<sup>13</sup> See the IALS User's Guide and Cascio, Clark and Gordon (2008) for further details.

<sup>14</sup> Holmlund, Lindhal and Plug (2011) argue that in countries such as Sweden, where compulsory school reforms were rolled out gradually across the country, mobility between municipalities is a potential threat to identification. In practice, however, they report that selective mobility has been a minor problem.

<sup>15</sup> We also exclude Eastern European countries because in the 1990s they were still in the early stages of their transition to market economies.

<sup>16</sup> This strategy is used – among others – by Brunello, Fort and Weber (2009) and Brunello, Fabbri and Fort (2013).

tests are closely correlated; b) for each value of prose literacy, there is an important variation in quantitative literacy. Table 1 shows the country means of years of schooling, prose and quantitative test scores in our final sample. We also include in the table a measure of average school quality, the number of teachers per 10 pupils in secondary education, which we take from Barro and Lee (2001). This measure is available every 5 years from 1960 to 1990 and is matched to our sample using birth cohorts.<sup>17</sup>

Table 2 documents the reforms affecting compulsory education which occurred in the European countries included in our sample from the 1930s until the late 1960s. For each reform, the table presents the year of the reform, the first birth cohort affected by the reform (or pivotal cohort), and the change in the minimum school leaving age. Years of compulsory schooling during the relevant sample period increased by one year in Great Britain, Ireland, the Netherlands and Northern Ireland and by three years in Denmark, Finland and Italy.<sup>18</sup>

#### 4. Results

As a preliminary step, we investigate whether Assumption 2 holds in our data. Since testing that  $\delta = 1$  and  $\varrho_1 \neq 0$  is equivalent to testing that  $\lambda_1 - \lambda_2 = 0$  and  $\theta_1 - \theta_2 = 0$  in equations (5a) and (5b), we estimate

$$T_{ij} = D_c + D_g + D_r + \lambda_j S_i + \theta_j Q_{cr} + \varepsilon_{ijr} \quad [12]$$

where  $j = 1, 2$  is for the test (prose or quantitative),  $i$  for the individual,  $c$  for the cohort,  $g$  for gender,  $r$  for the country,  $D$  is a set of cohort, gender and country dummies,  $S$  denotes years of schooling, and we control for gender, country and cohort dummies. Cohort dummies allow us to capture in a flexible way the underlying trend in test scores.<sup>19</sup> We estimate equation (12) by ordinary least squares. Since both  $Cov(S, \varepsilon)$  and  $Cov(Q, \varepsilon)$  are different from zero, these estimates of parameters  $\lambda_j$  and  $\theta_j$  are biased. However, under the reasonable assumptions that  $Cov(\varepsilon_1, S) = Cov(\varepsilon_2, S)$  and  $Cov(\varepsilon_1, Q) = Cov(\varepsilon_2, Q)$  – that is, that the correlation between years of schooling / school quality and unobserved ability does not vary with the type of test – the estimates of the differences  $\lambda_1 - \lambda_2$  and  $\theta_1 - \theta_2$  are unbiased.<sup>20</sup> By taking cross-test differences in the estimates of  $\lambda$  and  $\theta$ , we eliminate the bias itself, and can therefore verify whether the two tests we consider (prose and quantitative literacy) provide distinct information.

<sup>17</sup> To illustrate, we assign the 1980 value of the teacher-pupil ratio to the individuals born between 1962 and 1966, who were between age 14 and 18 in 1980.

<sup>18</sup> Denmark is the only country for which we consider two school reforms.

<sup>19</sup> We also experiment with a specification that adds to cohort dummies a set of country specific linear trends, with no qualitative change of results.

<sup>20</sup> See the Appendix for further details.

Table 3 presents both the estimates of the effects of years of schooling and school quality on test scores and the p-values associated to the test for each hypothesis.<sup>21</sup> We find that the hypothesis  $\lambda_1 - \lambda_2 = 0$  cannot be rejected at the five percent level of significance. However, we reject the hypothesis  $\theta_1 - \theta_2 = 0$  at the one percent level of significance.<sup>22</sup> These findings imply that the two tests provide distinct information, and that the test in equation (11) is equal to 1 under the null hypothesis of internal validity of the identification strategy.

Table 4 presents the results of the test, both when we regress prose literacy on quantitative literacy and when the dependent variable is quantitative literacy (the reverse regression). In each regression, we instrument the treatment variable with years of compulsory education.<sup>23</sup> We find that the estimated coefficient on prose (quantitative) literacy is statistically different from zero but not statistically different from one. This result suggests that the recent practice of estimating the causal effect of years of schooling using the exogenous variation produced by changes in compulsory education does not produce inconsistent estimates even if school quality is omitted from the regression.

## 5. Test Power

In this section, we use simulations to explore the power of the test to detect deviations from the null hypothesis. The data generating process is described by the following set of three equations

$$T_1 = \lambda_1 S + \theta_1 Q + \varepsilon_1 \quad [13]$$

$$T_2 = \lambda_2 S + \theta_2 Q + \varepsilon_2 \quad [14]$$

$$S = \varphi Z + \varepsilon_3 \quad [15]$$

where equations (13) and (14) reproduce (5) and (6) and equation (15) is the schooling equation.

To derive equation (15), assume that individuals maximize their utility net of schooling costs. Utility is log-linear in income. Schooling costs include opportunity, psychic and effort costs, and are described by the function  $C(S, Q, Z)$ , where  $Z$  is compulsory education and we assume  $C_S > 0$ ,  $C_{SS} > 0$ ,  $C_{SQ} \neq 0$ . There are different reasons why quality affects schooling costs: on the one hand, higher peer quality may facilitate learning and reduce individual effort. On the other hand, higher quality education may be more difficult to

<sup>21</sup> In each regression the standard errors are clustered by country and year of birth.

<sup>22</sup> Similar evidence is presented by Pekkala Kerr et al, 2013, who show that a Finnish reform affecting school quality had differential effects on reading and math test scores.

<sup>23</sup> In the first stage regressions, we find that years of compulsory education affect both quantitative and prose literacy (with coefficients equal to 0.010\*\* (st.err.: 0.005) and 0.011\*\* (st.err.: 0.005) respectively. We also find that years of compulsory education affect years of schooling (coefficient: 0.190\*\*\* and st.err.: 0.042).

access and more demanding to complete, which raises both effort and psychic costs. Using equation (3), where the outcome variable  $Y$  is log income, net utility maximization yields  $S = S(Q,Z)$ . A linear version of this equation is  $S = \varphi Z + \sigma Q + \eta$ , which corresponds to equation (15) if  $\varepsilon_3 = \sigma Q + \eta$ .<sup>24</sup>

Using the observed sample moments, we assume that the vector  $(Z, Q, \varepsilon_1, \varepsilon_2, \varepsilon_3)$  has a joint normal distribution with means [8.8, 0.8, 2, 2, 11.5] and set  $Var[Z] = 2.75$ ,  $Var[Q] = 0.05$ ,  $Var[\varepsilon_1] = 0.15$ ,  $Var[\varepsilon_2] = 0.2$ ,  $Var[\varepsilon_3] = 11$ ,  $Cov(Q, \varepsilon_1) = 0$ ,  $Cov(Q, \varepsilon_2) = 0$ ,  $Cov(Q, \varepsilon_3) = 0.0002$ ,  $Cov(\varepsilon_1, \varepsilon_2) = 0.15$ ,  $Cov(\varepsilon_1, \varepsilon_3) = 0.0004$ ,  $Cov(\varepsilon_2, \varepsilon_3) = 0.0002$ . In most simulations, we also use our estimates to set  $\lambda_1 = 0.062$ ,  $\lambda_2 = 0.061$ ,  $\theta_1 = 0.193$ ,  $\theta_2 = 0.094$  (see Table 3), and  $\varphi = 0.19$ . Rewriting (15) as  $S = \varphi Z + \sigma Q + \eta$  implies that  $Cov(Z, \varepsilon_3) = \sigma Cov(Z, Q) + Cov(Z, \eta)$ . For both the null hypothesis and the alternatives, we assume  $Cov(Z, \eta) = 0$ , which yields  $Cov(Z, \varepsilon_3) = \sigma Cov(Z, Q)$ . A regression of years of schooling on years of compulsory education and the teacher-pupil ratio in our data yields  $\sigma = 0.7$ . Finally, we assume that  $Cov(Z, \varepsilon_1) = 1.5 * Cov(Z, Q)$ ,  $Cov(Z, \varepsilon_2) = 1.4 * Cov(Z, Q)$ ,  $Cov(Z, \varepsilon_3) = 0.7 * Cov(Z, Q)$ .

We perform two different exercises: first, we check the power of the test  $(\frac{\beta_{IV}}{\gamma_{IV}} - 1)$  to detect deviations from the null hypothesis. To do so, we keep  $\lambda_1$  and  $\lambda_2$  as well as  $\theta_1$  and  $\theta_2$  fixed at their estimated values and allow  $Cov(Z, Q)$  to vary both between -0.05 and -0.25 and between 0.05 and 0.25. With the assigned values to the variances of  $Z$  and  $Q$ , this corresponds to allowing the correlation between  $Z$  and  $Q$  to vary within the ranges [-0.13/-0.68] and [+0.13/0.68]. Results are shown in Table 5.

Second, we check the sensitivity of the test statistic to deviations from Assumption 2 when  $Cov(Z, Q) = 0$ . Since this assumption can be violated either when  $\delta \neq 1$  or when  $\varrho_1 = 0$ , we consider both cases in turn. Setting  $Cov(Z, Q) = 0$ , we first vary  $\delta$ , i.e. the ratio of  $\lambda_2$  to  $\lambda_1$ , within the range [0.5, 1.5].<sup>25</sup> Since  $\varrho_1 = \theta_2 - \delta\theta_1$ , we keep  $\varrho_1$  at its estimated value (-0.099) when  $\delta$  varies by adjusting  $\theta_2$ . Results are reported in Table 6. Next, we set  $\delta$  equal to 1 and let  $\varrho_1$  approach zero from above or from below while keeping  $Cov(Z, Q) = 0$  (see Table 7).

In each table reporting our results (Table 5, 6 and 7) we show the mean, the standard deviation and the 95% confidence interval of the test statistic. In all simulations, we use a sample size of 5000 observations and 500 replications. Table 5 shows that the test is able to detect deviations from the null hypothesis –  $Cov(Z, Q) = 0$  – even when the covariance between  $Z$  and  $Q$  is as small as  $|0.05|$ . Figure 3 illustrates this result by presenting the empirical distribution of the test when we set  $Cov(Z, \varepsilon_1) = 0.15$ ,  $Cov(Z, \varepsilon_2) = 0.14$  and  $Cov(Z, \varepsilon_3) = 0.07$  and consider three different cases: a)  $H_0: Cov(Z, Q) = 0$ ; b)  $H_1^A: Cov(Z, Q) = 0.1$ ; c)  $H_1^B:$

<sup>24</sup> An alternative derivation of (16) assumes that costs are independent of quality and that log income is a function of the product of schooling  $S$  and quality  $Q$ , as in the Appendix.

<sup>25</sup> We set the value of  $\lambda_1$  at 0.0615 (the average between the estimated values of  $\lambda_1$  and  $\lambda_2$ ).

$Cov(Z,Q) = -0.1$ . Table 6 shows that it would take an implausibly large value of  $\delta$  – either below 0.7 or above 1.3 – to reject the null when  $Cov(Z,Q) = 0$ . Finally, Table 7 reports that allowing the value of  $\rho_t$  to approach zero does not lead to reject the hypothesis  $Cov(Z,Q) = 0$  when it is true.

## Conclusions

In this paper, we develop a new test for the validity of instrumental variables in models where nested measures of an endogenous variable are available. The test only requires two distinct indicators of a broader measure of the treatment variable and one instrument, and does not need information on outcomes. Its applicability is not restricted to settings in which both the instrument and the treatment are binary. Under the null hypothesis of instrument validity, the test statistic is computed by regressing one of the available broader measures of the treatment on the other using instrumental variables.

We apply this test to assess whether the identification strategy which relies on the variation provided by compulsory school reforms – a strategy extensively used in empirical labour economics – is a valid procedure to recover the causal impact of the quantity of schooling on outcomes such as earnings, health and crime when school quality is omitted because of lack of data. Using data on prose and quantitative literacy test scores, which depend both on the quantity and the quality of education, we cannot reject the hypothesis that compulsory school reforms are a valid instrument for years of schooling in regressions that omit school quality and other unobserved factors. We explore the performance of the test under the null and a set of alternative hypotheses using Monte-Carlo simulations that mimic the descriptive statistics of the variables in our data and find that the test has power to detect violations from the null hypothesis.

Our evidence supports the current widespread practice of using compulsory school reforms as sources of exogenous variation to estimate the causal effects of years of schooling on several economic outcomes, including earnings, crime and health.

## Appendix

### A.1. Nonlinear relationship between test scores, years of education and school quality

In this section we describe the test when the relationship between human capital, years of schooling and school quality is given by

$$H = \lambda S + \theta Q + \sigma SQ + \zeta \quad [A1]$$

Similarly, we assume that

$$T_1 = k_1 + \lambda_1 S + \theta_1 Q + \sigma_1 SQ + \zeta_1 \quad [A2]$$

$$T_2 = k_2 + \lambda_2 S + \theta_2 Q + \sigma_2 SQ + \zeta_2 \quad [A3]$$

In these specifications, the effect of years of schooling on human capital and test scores vary with school quality. In this case, the test statistic is given by

$$\frac{\beta_{IV}}{\gamma_{IV}} = \frac{\frac{Cov(Y, Z)}{Cov(T_1, Z)}}{\frac{Cov(Y, Z)}{Cov(T_2, Z)}} = \frac{Cov(T_2, Z)}{Cov(T_1, Z)} = \frac{\lambda_2 Cov(S, Z) + \theta_2 Cov(Z, Q) + \sigma_2 Cov(SQ, Z) + Cov(Z, \zeta_2)}{\lambda_1 Cov(S, Z) + \theta_1 Cov(Z, Q) + \sigma_1 Cov(SQ, Z) + Cov(Z, \zeta_1)} \quad [A4]$$

If  $S, Q$  and  $Z$  are jointly normal, then

$$Cov(SQ, Z) = E(Q)Cov(S, Z) + E(S)Cov(Q, Z)$$

(see, e.g., Bohrnstedt and Goldberger, 1969) and equation [A4] can be written as

$$\frac{\beta_{IV}}{\gamma_{IV}} = \frac{[\lambda_2 + \sigma_2 E(Q)]Cov(S, Z) + [\theta_2 + \sigma_2 E(S)]Cov(Z, Q) + Cov(Z, \zeta_2)}{[\lambda_1 + \sigma_1 E(Q)]Cov(S, Z) + [\theta_1 + \sigma_1 E(S)]Cov(Z, Q) + Cov(Z, \zeta_1)} \quad [A5]$$

under the alternative and

$$\frac{\beta_{IV}}{\gamma_{IV}} = \frac{[\lambda_2 + \sigma_2 E(Q)]}{[\lambda_1 + \sigma_1 E(Q)]} \quad [A6]$$

under the null.

In our data, we cannot reject the hypothesis that  $\frac{[\lambda_2 + \sigma_2 E(Q)]}{[\lambda_1 + \sigma_1 E(Q)]} = 1$  and that  $\theta_1 \neq \theta_2$ . Therefore,

Assumption 2 is still satisfied and the test is similar to the one carried out in the main text.

## A.2. The OLS bias in equations [5a] and [5b]

In this section we compute the bias associated to the OLS estimates of equations

$$T_1 = \lambda_1 S + \theta_1 Q + \varepsilon_1 \quad [\text{A.7}]$$

$$T_2 = \lambda_2 S + \theta_2 Q + \varepsilon_2 \quad [\text{A.8}]$$

Let  $S$  and  $Q$  be defined as deviations from means. Then the OLS estimates of  $\lambda$  and  $\theta$ , that we denote with an upper bar, are

$$\bar{\lambda}_1 = \lambda_1 + \frac{\text{Cov}(S, Q)}{\text{Var}(S)}(\theta_1 - \bar{\theta}_1) + \frac{\text{Cov}(S, \varepsilon_1)}{\text{Var}(S)} \quad [\text{A.9}]$$

$$\bar{\lambda}_2 = \lambda_2 + \frac{\text{Cov}(S, Q)}{\text{Var}(S)}(\theta_2 - \bar{\theta}_2) + \frac{\text{Cov}(S, \varepsilon_2)}{\text{Var}(S)} \quad [\text{A.10}]$$

$$\bar{\theta}_1 = \theta_1 + \frac{\text{Cov}(Q, \varepsilon_1)\text{Var}(S) - \text{Cov}(S, \varepsilon_1)\text{Cov}(S, Q)}{\text{Var}(Q)\text{Var}(S) - \text{Cov}(S, Q)^2} \quad [\text{A.11}]$$

$$\bar{\theta}_2 = \theta_2 + \frac{\text{Cov}(Q, \varepsilon_2)\text{Var}(S) - \text{Cov}(S, \varepsilon_2)\text{Cov}(S, Q)}{\text{Var}(Q)\text{Var}(S) - \text{Cov}(S, Q)^2} \quad [\text{A.12}]$$

Under the assumptions that

$$\text{Cov}(Q, \varepsilon_2) = \text{Cov}(Q, \varepsilon_1) \quad [\text{A.13}]$$

$$\text{Cov}(S, \varepsilon_2) = \text{Cov}(S, \varepsilon_1) \quad [\text{A.14}]$$

we obtain

$$\bar{\lambda}_1 - \bar{\lambda}_2 = \lambda_1 - \lambda_2 \quad [\text{A.15}]$$

$$\bar{\theta}_1 - \bar{\theta}_2 = \theta_1 - \theta_2 \quad [\text{A.16}]$$

Under the stated assumptions, the OLS estimates of the effects of  $S$  and  $Q$  on test scores are biased, but the difference of the estimates is unbiased.



## References

- Acemoglu Daron and Joshua Angrist (2001). How Large Are Human Capital Externalities? Evidence from Compulsory Schooling Laws. In *NBER Macroeconomics Annual 2000, Vol.15*, ed. Ben S. Bernanke and Kenneth Rogoff, 9-59. Cambridge, MA: National Bureau of Economic Research.
- Barro, Robert and Jong-Wha Lee (2001). Schooling Quality in a Cross Section of Countries. *Economica*, 68, 465-488.
- Bhuller, Manudeep, Magne Mogstad and Kjell G. Salvanes (2011). Life-Cycle Bias and the Returns to Schooling in Current and Lifetime Earnings. IZA Discussion Paper No. 5788.
- Black, Sandra E., Paul J. Devereux and Kjell G. Salvanes (2005). Why the Apple Doesn't Fall Far: Understanding Intergenerational Transmission of Human Capital. *American Economic Review*, 95(1), 437-449.
- Black, Sandra E. and Devereux, Paul J. (2011). Recent Developments in Intergenerational Mobility. In *Handbook of Labor Economics, Vol.4B*, ed. Orley C. Ashenfelter and David Card, 1487-1541. Amsterdam: Elsevier.
- Bohrnstedt, George W. and Arthur S. Goldberger (1969). On the Exact Covariance of Products of Random Variables. *Journal of the American Statistical Association*, 64, 1439-1442.
- Brunello, Giorgio and Daniele Checchi (2007). Does School Tracking Affect Equality of Opportunity? New International Evidence. *Economic Policy*, 22, 781-861.
- Brunello, Giorgio, Daniele Fabbri and Margherita Fort (2013). The Causal Effect of Schooling on Body Mass: Evidence from Europe. *Journal of Labor Economics*, 31(1), 195-233.
- Brunello, Giorgio, Margherita Fort and Guglielmo Weber (2009). Changes in Compulsory Schooling, Education and the Distribution of Wages in Europe. *Economic Journal*, 119, 516-539.
- Card, David (1999). The Causal Effect of Education on Earnings. In *Handbook of Labor Economics, Vol.3A*, ed. Orley C. Ashenfelter and David Card, 1801-1863. Amsterdam: Elsevier.
- Cascio, Elizabeth, Damon Clark and Nora Gordon (2008). Education and the Age Profile of Literacy into Adulthood. *Journal of Economic Perspectives*, 22(3), 47-70.
- Clark, Damon and Heather Royer (2013). The Effect of Education on Adult Health and Mortality: Evidence from Britain. *American Economic Review*, forthcoming.
- Durbin, James (1954). Errors in Variables. *Review of the International Statistical Institute*, 22, 23-32.
- Feinstein, Leon, Ricardo Sabates, Tashweka M. Anderson, Annik Sorhaindo and Cathie Hammond (2006). What Are the Effects of Education on Health? In *Report on Measuring the Effects of*

*Education and Health and Civic Engagement: Proceedings of the Copenhagen Symposium*, 171-354. Paris: Organisation for Economic Co-operation and Development.

Fort, Margherita, Nicole Schneeweis and Rudolf Winter-Ebmer (2011). More Schooling, More Children: Compulsory Schooling Reforms and Fertility in Europe. CEPR Discussion Paper No. 8609.

Garrouste, Christelle (2010). *100 Years of Educational Reforms in Europe: A Contextual Database*. EUR – Scientific and Technical Research series, Vol. 24487. Luxembourg: Publications Office of the European Union.

Green, David A. and W. Craig Riddell (2003). Literacy and Earnings: An Investigation of the Interaction of Cognitive and Unobserved Skills in Earnings Generation. *Labour Economics*, 10(2), 165-184.

Hanushek, Eric A. and Dennis D. Kimko (2000). Schooling, Labor-Force Quality, and the Growth of Nations. *American Economic Review*, 90(5), 1184-1208.

Hanushek, Eric A. and Ludger Woessmann (2012). Do Better Schools Lead to More Growth? Cognitive Skills, Economic Outcomes, and Causation. *Journal of Economic Growth*, 17(4), 267-321.

Hausman, Jerry A. (1978). Specification Tests in Econometrics. *Econometrica*, 46, 1251-1272.

Heckman, James J., Jora Stixrud and Sergio Urzua (2006). The Effects of Cognitive and Noncognitive Abilities on Labor Market Outcomes and Social Behavior. *Journal of Labor Economics*, 24(3), 411-482.

Holmlund, Helena and Mikael Lindahl and Erik Plut (2011) The Causal Effect of Parents' Schooling on Children's Schooling: A Comparison of Estimation Methods. *Journal of Economic Literature*, 49 (3), 615-651.

Huber, Martin (2013). Testing the Validity of the Sibling Sex Ratio Instrument. Department of Economics, University of St. Gallen, unpublished manuscript.

Huber, Martin and Giovanni Mellace (2013). Testing Instrument Validity for LATE Identification Based on Inequality Moment Constraints. Department of Economics, University of St. Gallen, unpublished manuscript.

Kitagawa, Toru (2008). A Bootstrap Test for Instrument Validity in the Heterogeneous Treatment Effect Model. Department of Economics, University College London, unpublished manuscript.

Lleras-Muney, Adriana (2005). The Relationship between Education and Adult Mortality in the United States. *Review of Economic Studies*, 72(1), 189-221.

Lochner, Lance (2011). Nonproduction Benefits of Education: Crime, Health, and Good Citizenship. In *Handbook of the Economics of Education, Vol. 4*, ed. Eric A. Hanushek, Stephen Machin and Ludger Woessmann, 183-282. Amsterdam: Elsevier.

Lochner, Lance and Enrico Moretti (2004). The Effect of Education on Crime: Evidence from Prison Inmates, Arrests, and Self-Reports. *American Economic Review*, 94(1), 155-189.

Machin, Stephen, Olivier Marie and Sunčica Vujić (2011). The Crime Reducing Effect of Education. *Economic Journal*, 121, 463-484.

Machin, Stephen, Kjell G. Salvanes and Panu Pelkonen (2012). Education and Mobility. *Journal of the European Economic Association*, 10(2), 417-450.

McCrary, Justin and Heather Royer. (2011). The Effect of Maternal Education on Fertility and Infant Health: Evidence From School Entry Policies Using Exact Date of Birth. *American Economic Review*, 101(1), 158-195

Meghir, Costas and Mårten Palme (2005). Educational Reform, Ability and Parental Background. *American Economic Review*, 95(1), 414-424.

Milligan, Kevin, Enrico Moretti and Philip Oreopoulos (2004). Does Education Improve Citizenship? Evidence from the U.S. and the U.K. *Journal of Public Economics*, 88(9-10), 1667-95.

Murnane, Richard J., John B. Willett and Frank Levy (1995). The Growing Importance of Cognitive Skills in Wage Determination. *Review of Economics and Statistics*, 77(2), 251-266.

Oreopoulos, Philip (2006). Estimating Average and Local Average Treatment Effects of Education when Compulsory Schooling Laws Really Matter. *American Economic Review*, 96(1), 152-175.

Pekkala Kerr, Sari, Tuomas Pekkarinen and Roope Uusitalo (2013). School Tracking and Development of Cognitive Skills. *Journal of Labor Economics*, 31(3), 577-602.

Sacerdote, Bruce (2011). Peer Effects in Education: How Might They Work, How Big Are They and How Much Do We Know Thus Far? In *Handbook of the Economics of Education, Vol.3*, ed. Eric A. Hanushek, Stephen Machin and Ludger Woessmann, 249-277. Amsterdam: Elsevier.

Sargan, John D. (1958). The Estimation of Economic Relationships using Instrumental Variables. *Econometrica*, 26, 393-415.

Wu, De-Min (1973). Alternative Tests of Independence Between Stochastic Regressors and Disturbances. *Econometrica*, 41, 733-750.

Figure 1. Quantitative literacy and years of schooling. International Adult Literacy Survey (IALS). Seven European countries (Denmark, Finland, Great Britain, Ireland, Italy, the Netherlands, Northern Ireland).

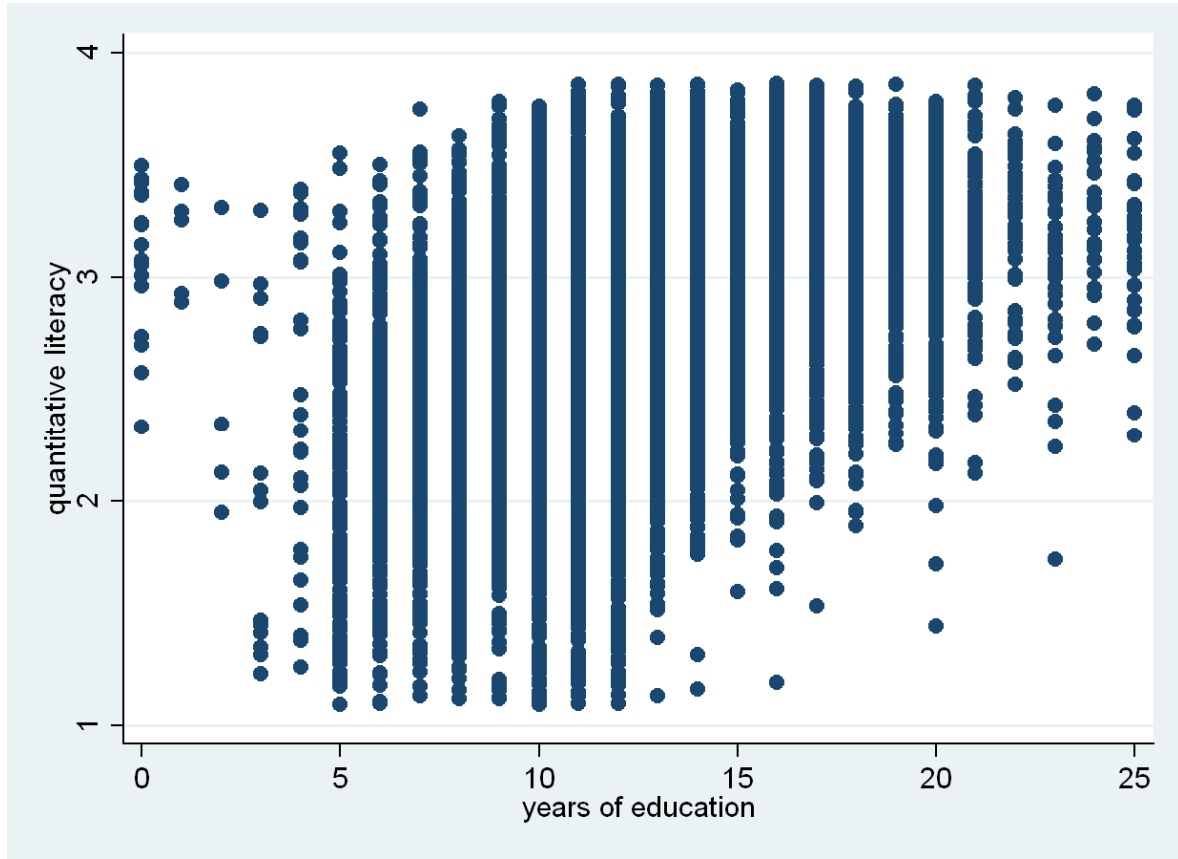


Figure 2. Quantitative and prose literacy. International Adult Literacy Survey (IALS). Seven European countries (Denmark, Finland, Great Britain, Ireland, Italy, the Netherlands, Northern Ireland).

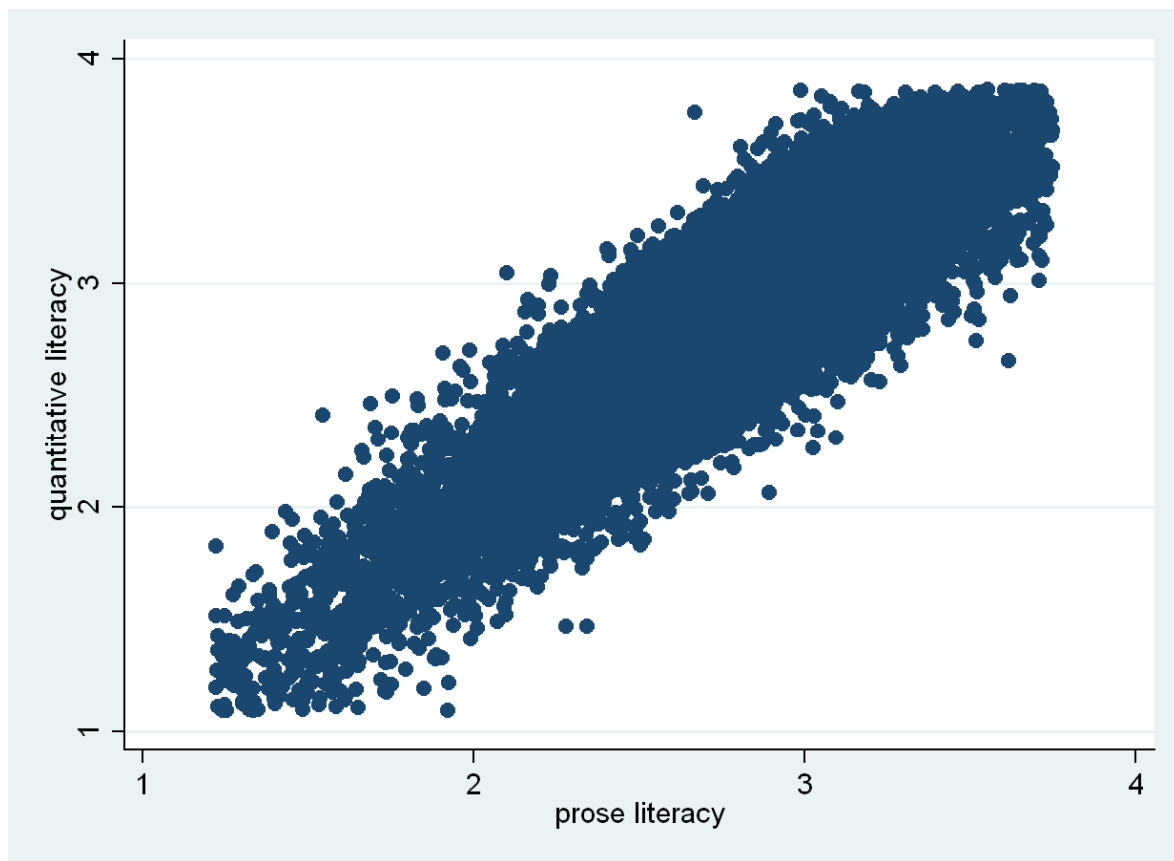
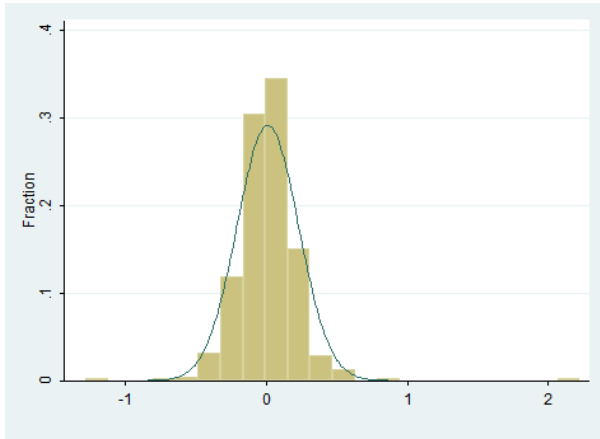
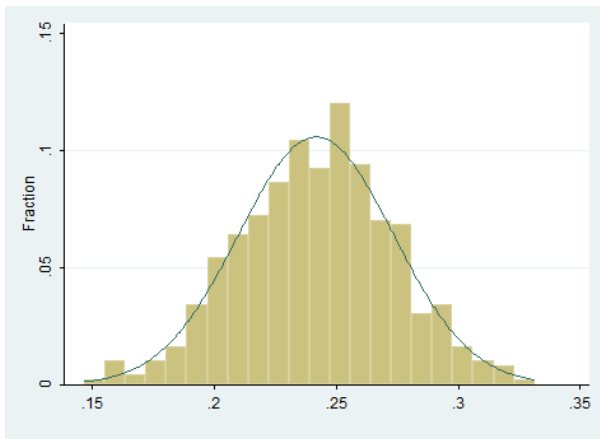


Figure 3. Empirical distribution of the test statistic under the null and two simulated deviations from the null hypothesis  $Cov(Z,Q) = 0$ . Simulated data with a sample size of 5000 observations and 500 replications

$$H_0 : Cov(Z,Q) = 0$$



$$H_1^A : Cov(Z,Q) = 0.1$$



$$H_1^B : Cov(Z,Q) = -0.1$$

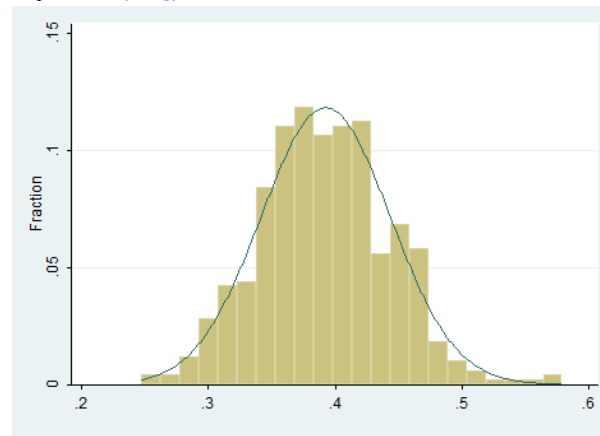


Table 1. Mean values of years of schooling, test scores and number of teachers per 10 pupils. Individuals born between 1942 and 1968.

	Years of schooling	Prose literacy score	Quantitative literacy score	Teachers per 10 pupils	Number of individuals
Denmark	13.15	2.78	3.01	0.928	2250
Finland	12.62	2.89	2.88	0.659	2107
Great Britain	12.44	2.76	2.76	0.642	2571
Ireland	10.22	2.67	2.65	0.700	1402
Italy	11.83	2.61	2.64	0.918	2164
Netherlands	13.26	2.90	2.94	0.601	2009
Northern Ireland	12.59	2.71	2.76	0.642	1961

*Notes:* Years of schooling and test scores are from IALS. The number of teachers per 10 pupils is taken from Barro and Lee (2001).

Table 2. Compulsory school reforms in Europe

	Reform year	Pivotal cohort	Change in minimum school leaving age	Years of compulsory education	Age at school entry
Denmark	1958	1947	11 to 14	4 to 7	7
–	1971	1957	14 to 16	7 to 9	7
Finland	1957	1951	13 to 16	6 to 9	7
Great Britain	1972	1958	15 to 16	10 to 11	5
Ireland	1966	1958	14 to 15	8 to 9	6
Italy	1963	1949	11 to 14	5 to 8	6
Netherlands	1968	1959	15 to 16	9 to 10	6
Northern Ireland	1957	1943	14 to 15	9 to 10	5

*Notes:* Data on school reforms are taken from Oreopoulos (2006), Garrouste (2010), Pekkala Kerr et al. (2012), Brunello et al. (2013) and Clark and Royer (2013).

Table 3. OLS estimates of literacy on years of schooling and school quality and p-values of the tests

<i>Prose literacy</i>	
Years of schooling	0.062*** (0.001)
Number of teachers per 10 pupils ratio	0.193*** (0.050)
<i>Quantitative literacy</i>	
Years of schooling	0.061*** (0.002)
Number of teachers per 10 pupils ratio	0.094* (0.054)
p-value of $\lambda_1 - \lambda_2 = 0$	0.085
p-value of $\theta_1 - \theta_2 = 0$	0.002
Observations	14464

Notes: Each regression includes gender, country and cohort dummies. Standard errors are clustered by country and cohort.

Table 4. IV estimates of prose literacy on quantitative literacy, and of quantitative literacy on prose literacy, and p-values of the test

<i>Prose literacy</i>	
Quantitative literacy	0.937*** (0.234)
$\frac{\beta_{IV}}{\gamma_{IV}} - 1 = 0$	0.789
<i>Quantitative literacy</i>	
Prose literacy	1.067*** (0.266)
$\frac{\beta_{IV}}{\gamma_{IV}} - 1 = 0$	0.802
Observations	14464

Notes: Each regression includes gender, country and cohort dummies. Standard errors are clustered by country and cohort.



Table 5. Power of the test statistic to detect violations of the assumption  $Cov(Z_i, Q) = 0$ . Simulated data with a sample size of 5000 observations and 500 replications

$Cov(Z_i, Q)$	Mean	S.D.	Quantile 0.025	Quantile 0.975
-0.25	0.376	0.024	0.328	0.42
-0.20	0.373	0.028	0.315	0.425
-0.15	0.375	0.034	0.304	0.437
-0.10	0.392	0.051	0.297	0.489
-0.05	0.508	0.149	0.274	0.831
0	0.013	0.218	-0.417	0.395
0.05	0.217	0.050	0.121	0.318
0.10	0.242	0.032	0.180	0.301
0.15	0.247	0.025	0.198	0.294
0.20	0.247	0.021	0.205	0.287
0.25	0.244	0.019	0.206	0.281

Table 6. Sensitivity of the test statistic to violations of Assumption 2. Parameter  $\delta$ , the ratio of  $\lambda_2$  to  $\lambda_1$ , is allowed to vary in the range  $[0.5, 1.5]$ , while  $\rho_1 = -0.099$  and  $Cov(Z_i, Q) = 0$ . Simulated data with a sample size of 5000 observations and 500 replications

$\delta$	Mean	S.D.	Quantile 0.025	Quantile 0.975
0.50	-0.709	2.967	-1.459	-0.140
0.60	-0.574	2.512	-1.278	-0.040
0.70	-0.439	2.057	-1.073	0.056
0.80	-0.304	1.605	-0.818	0.165
0.90	-0.169	1.156	-0.634	0.279
1.00	-0.034	0.719	-0.439	0.442
1.10	0.101	0.341	-0.303	0.660
1.20	0.237	0.367	-0.182	0.879
1.30	0.372	0.757	-0.652	1.087
1.40	0.507	1.196	0.052	1.297
1.50	0.642	1.645	0.145	1.546

Table 7. Sensitivity of the test statistic to violations of Assumption 2. Parameter  $\varrho_1$  approaches 0 from above or below while  $\delta = 1$  and  $Cov(Z, Q) = 0$ . Simulated data with a sample size of 5000 observations and 500 replications.

$\varrho_1$	Mean	S.D.	Quantile 0.025	Quantile 0.975
-0.100	-0.034	0.720	-0.440	0.442
-0.075	-0.032	0.691	-0.450	0.443
-0.050	-0.031	0.661	-0.453	0.443
-0.025	-0.029	0.633	-0.455	0.444
0.000	-0.027	0.604	-0.457	0.445
0.025	-0.026	0.577	-0.459	0.446
0.050	-0.025	0.549	-0.461	0.447
0.075	-0.023	0.523	-0.463	0.450
0.100	-0.022	0.498	-0.466	0.455