KIER DISCUSSION PAPER SERIES

KYOTO INSTITUTE OF ECONOMIC RESEARCH

http://www.kier.kyoto-u.ac.jp/index.html

Discussion Paper No. 630

"Does Secondary School Tracking Affect Performance? Evidence from IALS"

by

Kenn Ariga

and

Giorgio Brunello

February 2007

KYOTO UNIVERSITY

KYOTO, JAPAN

Does Secondary School Tracking Affect Performance? Evidence from IALS^

by

Kenn Ariga (Kyoto University)

Giorgio Brunello* (University of Padova, CESifo and IZA)

^ This paper was produced while the second author was visiting Kyoto University, which provided excellent hospitality.

* Corresponding author: Giorgio Brunello, Department of Economics, University of Padova, via del Santo 33, 35100 Padova, Italy. E-mail: giorgio.brunello@unipd.it

Abstract

There is substantial cross - country variation in secondary school design, with some countries tracking students into different ability schools very early, and other countries with little or no tracking at all. Does tracking length affects school performance, as measured by standardized test scores? We use the international data from the International Adult Literacy Survey to estimate the relationship between the experienced tracking length and the performance in standardized cognitive test scores of young adults, aged between 16 and the mid - twenties. Our IV estimates suggest that the contribution of tracking to performance is positive and statistically significant: conditional on total years of schooling, one additional year spent in a track raises average performance by 3.3 to 3.4 percentage points, depending on the estimates.

Key words: tracking, secondary schools, efficiency

JEL codes: I21, I28

Introduction

In many education systems around the world, heterogeneous pupils are initially mixed in comprehensive schools - typically primary and lower secondary education. At some stage of the curriculum, some form of (self) selection takes place, typically based on ability and past performance, and students are allocated to schools which specialize in different curricula (tracks) or to classes where subjects are taught at a different level of difficulty (streams). The former system is typical of Central European countries, such as Germany, Austria, The Netherlands and Hungary, but also of Korea, China, Brazil, The Russian Federation, Egypt and Japan, and the latter system is typically observed in the US.

We define the allocation of students into different schools or classes as "tracking". Selection and diversification of curricula are typical features of college education (see Shavitt et al, 2007), but in this paper we are concerned exclusively with the tracking of secondary schools. Critics of early tracking systems argue that high-performing students gain from tracking at the expense of their lower-performance peers. While well placed, this concern with equality of opportunity prompts the following question: is there a trade-off between equality of opportunity and efficiency, and if yes, what are the efficiency costs of de-tracking secondary schools? A natural way of measuring the relative efficiency of tracking is to investigate whether selective schooling affects individual accumulation of human capital, as measured by the results of standardized cognitive tests. Such investigation is informative for the question at hand if productivity at school and in the labour market are correlated, and the costs of tracked and untracked systems are roughly comparable. Even granting this, the empirical task is fraught with difficulties, as documented for the US by Betts and Shkolnik, 2000, and Figlio and Page, 2000, among others.

In a recent study, Hanushek and Wossmann, 2006, (HW from now on), adopt a difference in difference strategy in a multi-country setup, and compare the standardized test scores of primary school students, who are typically not tracked in all countries, with those of older students, aged 14 to 15, who in some countries – most notably Germany, Austria and Hungary - are already in a tracked school. Their strategy has two merits: it ensures that the allocation to treatment (tracking) is exogenous, and it generates the within - country variation required to evaluate the impact of a country – specific institution such as tracking. HW find evidence that early tracking increases inequality, but no evidence that it positively affect average school performance. If anything, performance declines in countries with earlier tracking.

Since the treatment group in the HW sample is very young, one potential problem with this interesting study is that, unless tracking has almost immediate effects, it cannot properly evaluate the contribution of tracking to performance in the countries where tracking starts at 15 or later[1]. According to the OECD, the age of first selection into tracking or streaming is either 15 or 16 in the majority of OECD countries, including the US, the UK, France, Japan and Scandinavia, to which one should add large non OECD countries, such as China, the Russian Federation, Brazil and Egypt. Since the early tracking analyzed by HW is likely to be mainly that of central continental Europe (Germany and some of its neighbours), a natural question is whether their results hold in a larger sample of countries.

In this paper, we investigate the efficiency of secondary school tracking using a different empirical strategy. We share with HW the use of a multi-country setup, but focus instead on a sample of older individuals – aged between 16 and the mid twenties, depending on the country – who have experienced in their own country different *tracking lengths*. Therefore, instead of comparing individuals who have been exposed to tracking to individuals without such exposure, we compare individuals with different degrees of exposure to the same policy, including no exposure at all. For this purpose, we use the International Adult Literacy Survey (IALS), an international survey of adults aged 16 or older sponsored by the OECD and covering 19 countries[2].

By focusing on older individuals, our approach has the advantage that the investigation is not restricted to early tracking countries. Moreover, by comparing individuals of different age within the same country, we can ask whether school performance is affected by tracking length, rather than by the mere existence of a tracking system. One disadvantage of our approach is that assignment to treatment cannot be considered as exogenous, because the variation of tracking length in our sample is partly due to age and country effects, and partly the result of individual behaviour, which affects the likelihood of dropping out of school.

We use instrumental variables to deal with endogenous selection into tracking, and find that one additional year of school spent in a track contributes significantly more to school outcomes than the same year spent in a comprehensive system. We hasten to stress that our estimates uncover the effect of tracking on school performance for the sub-sample of

[1] Because the average student age in the PISA dataset is 15 years and 9 months, the authors claim that they can also capture the effects of tracking in systems that track at 15. This is difficult to believe if the effect of tracking on school performance operates via peer effects of differential learning.

[2] These countries are: Switzerland, Germany, the US, the UK, Ireland, The Netherlands, Poland, Sweden, New Zealand, Belgium, Italy, Norway, Slovenia, The Czech Republic, Denmark, Finland, Hungary, Canada and Chile.

individuals whose treatment status is changed by the selected instrument (compliers), and do not automatically apply to a randomly chosen individual. It is well known that local average treatment effects can differ substantially from average treatment effects if the fraction of the population affected by the instruments is small (Oreopoulos, 2006). Still, we believe that our results cast some doubt on the view that de-tracking secondary schools have positive equity implications without any efficiency cost.

The paper is organized as follows: in the next section we describe in detail our empirical strategy. Section 2 introduces the data and Section 3 presents the results. A final brief section concludes.

1. The Empirical Strategy

Consider the following empirical model

$$A_{ic} = \alpha_c + X_{ic}\beta + \gamma E_{ic} + \varepsilon_{ic} \qquad [1]$$

where the subscripts i and c are for the individual and the country, A is the log of individual achievement in a standardized cognitive test, α is a country specific intercept, X is a vector of individual controls, and E is years of completed education from primary to upper secondary school, net of repeated grades[3]. This specification implies that the contribution of each year of education to the log score is constant and equal to γ.

Let T be the number of years spent in a tracking system, be it academic / vocational as in the European or Asian context, or honours/general as in the US (see Rees, Argys and Brewer, 1996)[4]. Conditional on the number of years of education, time spent in tracking improves log performance if

$$A_{ic} = \alpha_c + X_{ic}\beta + \gamma E_{ic} + (\sigma - \gamma)T_{ic} + \varepsilon_{ic} \qquad [2]$$

where $(\sigma - \gamma) > 0$. In this specification, one year of education in a comprehensive system and in a tracking system add γ and $\sigma > \gamma$ to log performance respectively. Equation [2] can also be written more explicitly as

[3] By excluding repeated grades, we assume that repetition does not contribute by itself to performance in a test.
[4] We are grateful to Dominic Brewer for advice on the US institutional setup.

$$A_{ic} = \alpha_c + X_{ic}\beta + \gamma NT_{ic} + \sigma T_{ic} + \varepsilon_{ic} \quad [3]$$

where NT is the number of years spent in a comprehensive system, and $E=NT+T$.

As discussed in the related literature, a number of problems need to be overcome to estimate equations [2] or [3]. First, school design, including tracking, is typically country - specific. If all students in a country are subject to the same system, the impact of tracking cannot be disentangled from the host of other country - specific effects on school performance. More precisely, let $T_{ic} = T_c$. If we control country - specific effects with country dummies, the effect of tracking is not identified, because of the lack of a control group. HW cleverly address this problem by noticing that no country tracks individuals in primary schools. Therefore, pupils in these schools can be used as the control group. They combine survey data which include test scores from primary and secondary schools and adopt a difference in difference approach.

To identify the impact of tracking, we focus on an alternative source of within - country variation, generated by the fact that, even within the same country, pupils are exposed to different tracking lengths, either because of age – with older pupils having more tracking than younger pupils – or because, conditional on country and age, some students drop out before completing upper secondary education. To illustrate, let Z_s and Z_e be the years of education when tracking starts and upper secondary education is expected to be completed, respectively. To exemplify, in Germany Z_s and Z_e are equal to 4 and 13. Then individual tracking length in secondary schools is given by

$$\begin{aligned} &T = 0 && \text{if} && Z \leq Z_s \\ &T = E - Z_s && \text{if} && Z_s < E < Z_e \end{aligned} \quad [4]$$

According to definition [4], within - country variation in tracking length is guaranteed by a) age variation, as younger individuals have not yet completed upper secondary education and the entire tracking span; b) educational variation for a given age, as some individuals drop out of school before completing a degree; c) curriculum differences, because in some countries individuals enrolled in a vocational track can typically complete education one year earlier than pupils in the general track – see for instance Germany and Switzerland – and thereby be exposed to shorter tracking.

A second problem with the estimation of [3] is the presence of unobserved individual characteristics, such as ability or motivation. If these unobservables are correlated with the education variables *NT* and *T*, because less academically talented or less motivated individuals are more likely to drop out of school before completion of upper secondary education, ordinary least squares estimates of [3] are biased (see the discussion in Betts and Shkolnik, 2000). As discussed by Heckman and Li, 2002, there are three main strategies to attenuate ability bias: a) instrumental variables; b) fixed effect estimators; c) selection on observables, which consists of including in the regression a vector of variables that proxy unobserved ability, motivation or effort. The second approach requires panel data. In the absence of such data, our empirical strategy combines points a) and c).

We include in [3] a large number of observables, which capture a) family background: parental education, occupation, and country of birth; b) motivation: dummy variables for early withdrawal from school due to boredom, dislike of school, and desire to learn a trade; c) health: dummy variables for visual, hearing, learning and speech disabilities, both current and at the time of primary and secondary school; d) current cultural and social activities, such as the use of public libraries, attending a cultural or sporting event, volunteering, reading books and newspapers and listening to music; e) self assessed literacy: dummy variables for individual rating of writing and reading skills, and indicators of dependency on others for basic literacy tasks; f) work activities: labour market status and type of work.

As shown in Table 1, the end of compulsory education typically occurs either at the age when tracking starts or later. Therefore, the years *NT* spent in a comprehensive school are also years of compulsory education. While the implementation of compulsory schooling laws is not always perfect, OECD enrolment data for 2004 show that more than 90 percent of the relevant age cohort is still in school before the end of compulsory education in most of the countries included in the table. This percentage falls substantially after the end of compulsory education, at least outside of Scandinavia, with the decline ranging from 11 percentage points in Germany to 33 percentage points in the UK and 35 percentage points in the US (OECD, 2006, Table C1.3)[5]. These facts motivate our assumption that, conditional on the set of observables described above, the number of years *NT* spent in comprehensive schooling are exogenous, and driven mainly by compulsory schooling laws. Given this assumption, which

[5] We are implicitly assuming that the 2004 figures are relevant also for schooling done 10 years earlier.

we test in the data using the standard Durbin-Wu-Hausman endogeneity test, equation [3] contains a single endogenous variable, tracking length *T*. We address this problem by using instrumental variables.

2. The Data

The IALS survey was implemented in three different waves – 1994, 1996, 1998 – using a common questionnaire, with the purpose of collecting comparable information on adult literacy. Literacy in the survey 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 tip, completing an order form, or determining the amount of interest on a loan[6]. Results of the tests are scaled in the range (0,500). We use the log of the simple average of the three available measures of literacy as the dependent variable in [3].

We select the sub-sample of young individuals aged between 16 and age *t-1970*, where *t* is the year of the country – specific survey, which ranges from 1994 to 1998[7]. There are three reasons for this choice: first, we wish to reduce the recall error in the number of years of completed education, which is likely to be lower the closer the individual is to the time spent in formal education. Second, we expect the contribution of schooling to the literacy test scores of young individuals to be predominant when compared to older adults, who are more likely to be affected in their scores by sub-sequent labour market experience. Last but not least, since the IALS survey has been taken between 1994 and 1998, depending on the country, our selected sample has experienced secondary school between the early 1980s and the second part of the 1990s. By choosing this age window, we reduce the risk of including individuals who have been exposed to different secondary school systems within the same country, due to educational reforms. This risk leads us to exclude two countries from our data: the Czech

---

[6] See the IALS User's Guide for further details.

[7] Since we want to focus on secondary education, we exclude from the sample those individuals who have completed a number of years of education – net of repetition - higher than Z_e or have attained a level of education higher than ISCED 3, which corresponds to upper secondary education.

Republic, which anticipated tracking from 15 to 11 in the early 1990s by allowing pupils to complete their compulsory education either in normal school or in the more selective gymnasium, and The Netherlands, which introduced in 1993 an additional year of compulsory education, thereby delaying tracking to age 13 (see Brunello and Checchi, 2007)[8].

The construction of individual tracking length requires information on the year when primary education starts, upper secondary education finishes, and tracking starts. This information is reported in Table 1. We use several sources to collect the relevant data, including OECD *Education at a Glance*, various issues, the UNESCO website (http://www.unesco.org), the Eurydice website (http://www.eurydice.org), national sources and Table 1 in Brunello and Checchi, 2007, which reports the age of first selection into tracks in the mid 1980s and mid 1990s[9]. Finally, the information on the years of completed education is taken from the following question in the IALS questionnaire: *"During your lifetime, how many years of formal education have you completed, beginning with first level, first grade, first stage, and not counting repeated years at the same level?"*

3. Results

Standard human capital theory suggests that rational individuals should drop out of school when the expected benefits from continuing education are inferior to the expected costs. Some of these benefits and costs depend on individual ability, motivation and effort, which are also correlated with school performance. Other costs depend instead on external constraints – such as lack of financial resources, illness or family reasons – and are not correlated in any obvious way to ability or motivation. Therefore, empirical measures of these costs are good candidates as instrumental variables for the years of tracking.

The IALS survey includes a question asking the main reason why individuals stopped their schooling when they did. This question is addressed to individuals who have not yet completed upper secondary or higher education, and picks up the reasons for dropping out of school beforehand. We use the replies to construct three dummy variables: a) a dummy taking the value 1 if the individual stopped school because of financial constraints, and zero otherwise; b) a dummy taking the value 1 if the individual stopped school because of family

[8] We also exclude Canada because of the limited information on parental background.
[9] While streaming in the US can start early depending on the school, we select the 10th grade as the age of first selection, in line with the OECD and based on the statistical information provided by Rees, Argys and Brewer (1996).

reasons, such as marriage, pregnancy and illness at home, and zero otherwise; c) a dummy taking the value 1 if the individual stopped school because of personal constraints (illness) or institutional constraints (school not available/not accessible). We use these three dummies as instruments of tracking length, and test for instrument validity using the Hansen J test.

Table 2 presents the means and standard deviations of the key variables used in the regressions. The average test score in our sample of 6984 individuals is 267.08 (standard deviation: 53.05), and the average number of years of education is 10.70, of which 8.13 are spent in a comprehensive schooling system. Close to 9 percent of the individuals in our sample has been affected by financial, family, health and school problems, leading to premature interruption of school. Importantly, dropouts because of exogenous reasons differ from the remaining sample of dropouts only marginally in the number of years spent in a comprehensive school (8.11 versus 8.22) and more significantly in the number of years spent in a tracked school (1.06 versus 1.60).

We estimate two alternative specifications of equation [3]: in the first and more parsimonious specification, we capture country specific effects with country dummies and age effects with a second order polynomial in age. In the second, we use country by age dummies. Since the number of observations varies by country, we use weighted regressions so as to give each country the same weight in the final sample. Unobserved heteroskedasticity is treated using robust standard errors. We test for the endogeneity of NT and T using the Durbin – Wu – Hausman test and find that the null hypothesis of no endogeneity is rejected for tracking years T (p-value of the test: 0.000) but not for the years NT in a comprehensive school (p-value: 0.482). These results confirm our maintained hypothesis that NT can be treated as exogenous.

Our key results are presented in Table 3. The table is organized in four columns. The first two columns are reserved to the more parsimonious specification, and the latter two columns to the specification with country by age dummies. Moreover, we present OLS estimates in the odd columns and IV estimates in the even columns. For simplicity, we only report the coefficients associated to the education variables. The estimates of the other coefficients are available from the authors upon request. We notice that the Bound F test of excluded instruments cannot reject the null hypothesis (joint significance), which signals that the selected instruments are not weak. Moreover, the Hansen J test for the validity of instruments always rejects the null of misspecification. Our OLS estimates (columns (1) and (3)) show that the years spent in a comprehensive school pay off in terms of standardized test scores about as much as the years spent in a tracked school, which suggests that allocation to tracking does not raise school performance as measured by our selected indicator.

On the other hand, our IV estimates (columns (2) and (4)) tell a different story: we find that one additional year of tracking raises the attained score by 4.5 to 4.6 percent, compared to the 1.1 to 1.2 percent induced by an additional year in a comprehensive school, depending on the specification. Furthermore, the difference in returns is statistically significant. These estimates suggest that spending one additional year in a tracked school rather than in a comprehensive school raises performance by 3.3 to 3.4 percent.

The finding that IV estimates of the returns to tracking are higher than OLS estimates can be due to a variety of reasons, including the fact that constrained individuals who are forced to drop out have higher returns to schooling than the average individual. Under the maintained assumption that differences in school outcomes are positively correlated to differences in productivity, and that the cost of tracking and non tracking systems are not substantially different, our findings point out that additional tracking in secondary school contributes to higher productivity and efficiency[10].

For the sake of illustration, suppose that our results are general enough to be applied to the relevant population, and that we can compare two individuals with the same observable characteristics but different type of schooling: the first student has 10 years of comprehensive schooling and 2 years of tracking, and the latter student has 4 years of comprehensive education and 8 years of tracking. Then our estimates predict that the performance of the latter student in standardized test scores is about 20 higher (6 x 0.033) than the performance of the former. This is a significant effect, equivalent to about 1 standard deviation of the log test score.

How general are our results? As discussed by Angrist, 2003, IV estimates capture a local average treatment effect (LATE), that is, the effect of the instruments on the sub-population of compliers, who have changed their behaviour as a result of the treatment. LATE estimates have two important features: a) they can differ substantially from the average effect of additional tracking on a randomly selected individual (ATE), especially when the fraction of the population affected is small. In our sample, the percentage of dropouts because of credit constraints, illness and family reasons is 9.83 percent; b) they are instrument dependent. Therefore, our answer to the question on the generality of our results has to be negative. Even

[10] The fact that our IV estimates of the contribution of tracking to performance are significantly higher than the OLS estimates is in line with the returns to education literature, which typically finds that estimated returns based on IV are much larger than OLS estimates.

so, we do believe that our results cast some doubt on the view that de-tracking secondary school is a “free lunch”, which buys additional equality at no expense in terms of efficiency.

Conclusions

The question addressed in this paper is whether tracking length in secondary schools affects school performance, as measured by standardized test scores. We have used the international data from the International Adult Literacy Survey to estimate the relationship between tracking length and the performance in standardized cognitive test scores of young adults, aged between 16 and the mid - twenties. Our IV estimates suggest that the contribution of tracking to performance is positive and statistically significant: conditional on total years of schooling, one additional year spent in a track raises average performance by 3.3 to 3.4 percentage points, depending on the estimates. While we do not claim that these findings are general, they suggest that views that de-tracking secondary schools generate equity gains with no efficiency losses may be too optimistic.

Table 1. School Design in a sample of OECD Countries.

	Age when compulsory education ends	*Age when tracking starts (1980-1990)*	*Age when secondary school finishes*	*Age when primary education starts*
Switzerland	15	15	20	7
Germany	18	10	19	6
Usa	17	16	18	6
Ireland	16	15	18	6
Poland	15	15	19	7
Sweden	16	16	20	7
New Zealand	16	16	18	6
UK	16	16	18	5
Belgium	18	12	19	6
Italy	14	14	19	6
Norway	16	16	19	6
Slovenia	15	15	18	6
Denmark	16	16	20	7
Finland	16	16	19	7
Hungary	16	10	20	6
Chile	14	14	18	6

Source: see text.

Table 2. Summary Statistics of the key variables

	Sample Mean	*Standard Deviation*
Average test score	267.59	53.03
Log average test score	5.563	0.232
Gender	0.487	
Age	20.589	3.493
Foreign born	0.042	
Single	0.676	
Lives in urban area	0.608	
Years of education	10.702	1.735
Years in comprehensive school	8.128	1.944
Year in tracking	2.536	2.259
Dummy: dropped out because of: financial reasons	0.048	
Dummy: dropped out because of: family reasons	0.041	
Dummy: dropped out because of: illness or no school available	0.009	

Table 3. Effects of tracking length of test performance. Dependent variable: log average test score in prose, documentation and quantitative knowledge.

	(1)	*(2)*	*(3)*	*(4)*
Years in comprehensive school	.020***	.012**	.019***	.011**
	(.005)	(.005)	(.003)	(.005)
Years in tracked school	.022***	.045***	.023***	.046***
	(.002)	(.006)	(.002)	(.007)
σ-γ	.002	.033***	.004	.034***
	(.005)	(.010)	(.004)	(.010)
Country Dummies	Yes	Yes	No	No
Country by Age Dummies	No	No	Yes	Yes
Family Background Controls	Yes	Yes	Yes	Yes
Health Controls	Yes	Yes	Yes	Yes
Controls for Self-Assessed Literacy	Yes	Yes	Yes	Yes
Controls for Cultural and Social Activities	Yes	Yes	Yes	Yes
Controls for Labour Market Status	Yes	Yes	Yes	Yes
Controls for Motivation and Attitude toward School	Yes	Yes	Yes	Yes
R Squared	.492	.480	.494	.483
Number of observations	7221	7221	7221	7221
Bound F test		293.93		279.33
		(.000)		(.000)
Hansen J test (2 degree of freedoms)		0.469		0.770
		(.791)		(.680)

Note: robust standard errors within parentheses. Each regression includes a constant, dummies for gender, urban residence and marital status. Columns (1) and (3) include also individual age and its square.

References

Angrist, J., 2003, Treatment Effect Heterogeneity in Theory and Practice, *The Economic Journal*, 114, C52-C83.

Betts, J. and Shkolnik, J., 2000, The Effects of Ability Grouping on Student Achievement and resource Allocation in Secondary Schools, *Economics of Education Review*, 19, 1-15

Brunello, G. and Checchi, D., 2007, Does Tracking Affect Equality of Opportunity? New International Evidence, forthcoming in *Economic Policy*.

Figlio, D. and Page, M., 2002, School Choice and the Distributional Effects of Ability Tracking: does Separation Increase Inequality? *Journal of Urban Economics*, 51, 3, 497-514.

Hanushek, E., and Woessmann, L., 2006, Does Education Tracking Affect Performance and Inequality? Differences in Differences Evidence Across Countries, *The Economic Journal*, 116, C63-C76.

Manning, A. and J.S.Pisckhe. 2006. Comprehensive versus Selective Schooling in England in Wales: What Do We Know? *IZA Discussion Paper n.2072*.

Minter Hoxby, C. 2000. Peer effect in the classroom: learning from gender and race variation. *NBER wp.7867*

Oakes, J. 1992. Can Tracking Research Inform Practice? Technical, Normative, and Political Considerations. *Educational Researcher* 21(4): 12-21

Oreopoulos, P., 2006, Estimating Average and Local Average Treatment Effects of Education when Compulsory Schooling Laws Really Matter, *American Economic Review*, 152-75

Rees, D., Argys, L., Brewer, D., 1996, Tracking in the United States: Descriptive Statistics from NELS, *The Economics of Education Review*, 83-89

Shavit, Y., R. Arum, and A. Gamoran with G. Menahem, (eds.) 2007, *Stratification in Higher Education: A Comparative Study*, Palo Alto: Stanford University Press.