

Kristian Koerselman
Anticipatory effects of curriculum
tracking

Aboa Centre for Economics

Discussion Paper No. 47
Turku 2009



Copyright © Kristian Koerselman

ISSN 1796-3133

Printed in Uniprint
Turku 2009

Kristian Koerselman
Anticipatory effects of curriculum tracking
Aboa Centre for Economics
Discussion Paper No. 47
May 2009

ABSTRACT

Curriculum tracking, the separation of secondary school students into academic and vocational tracks, correlates positively with pre-tracking achievement in both British and international data. I argue that this correlation is caused by the incentives emanating from the track placement decision. Using test score data collected in TIMSS 1995 and 2003, and in PIRLS 2001 and 2006, I investigate the effect of tracking on the early achievement distribution empirically, amongst others by means of quantile regression. The evidence presented in this paper implicates that previous value-added estimates of the net impact of tracking may be biased.

JEL I21, I28, J08, J24

Keywords: curriculum tracking, ability streaming, anticipatory effects, high-stakes testing

Contact information

Kristian Koerselman
Department of Economics and Statistics
Åbo Akademi University
20500 Turku
Finland.
tel.: +358 2 215 4329
fax: +358 2 215 4677
koerselman@economistatwork.com
<http://economistatwork.com>

Acknowledgements

I thank Tuomas Pekkarinen, Markus Jäntti, Ludger Woessmann, Heikki Kauppi, Jonas Lagerström, Sari Kerr, Roope Uusitalo and Elias Einiö for their kind help and advice. I gratefully acknowledge financial support from Yrjö Jahnessonin säätiö, Stiftelsens för Åbo Akademi forskningsinstitut, Bröderna Lars och Ernst Krogius forskningsfond, and from the Academy of Finland.

1 Introduction

Curriculum tracking is the the separation of secondary school students into academic and vocational tracks according to expected achievement. Whether tracking is preferable is open to debate. Some argue that students learn better when receiving education adapted to their level. Furthermore, students may need to specialize in certain types of knowledge; the demand for vocational skills may be limited in academic professions and vice versa.

On the other hand, it can be argued that tracking forces career paths onto students at too early an age. Intergenerational mobility may suffer if young children's choices are effectively determined by their parents while older ones can think for themselves. A related argument is that 'expected achievement' is measured with significant noise at a younger age. This could lead to a suboptimal allocation of individuals to professions.

We do not only want to maximize the expected educational achievement of students: we are interested in the *distribution* of achievement too, if only because it influences the future distribution of earnings. Curriculum tracking plays an important role in determining the variance of achievement. This is because a student's achievement tends to regress to the mean achievement of his classmates through so-called *peer effects*. Credible estimates of the size of peer effects are as large as between 11% and 40% of inter-student variation (Hoxby, 2000, Ammermueller and Pischke, 2006). Tracking explicitly affects class composition, and through the peer effect the distribution of achievement.

It is intuitive to think that tracking should lead to larger population-wide differences between students: when classes are more homogeneous, aggregate peer effects should be smaller. Empirical studies tend to confirm this. Even so, the sign of the effect on average scores is still an open question.

The effect of tracking may extend to the grades before its start. Students have an incentive to work harder before the selection point in order to end up in a higher track. In this paper, I therefore argue that tracking should have a positive effect on pre-tracking scores. I go on to investigate the matter

empirically, and find some evidence for such *anticipatory effects*.

If it is indeed the case that anticipatory effects exist, there are implications for the existing literature on curriculum tracking. A number of studies on tracking uses value-added models to control for omitted variables. These models operate under the assumption that no anticipatory effects exist, and will lead to biased estimates if the opposite is the case.

2 Background

Many countries have traditionally tracked students from an age of 12 or so onward, but during the last five decades, a number of them has postponed tracking to the middle or even to the end of secondary education. This has provided us with natural experiments on tracking policies. Meghir and Palme (2005), Pekkarinen (2005), and Pekkarinen et al. (2007) use variation between subsequent cohorts at the time of the reform. In some cases, the reform was not implemented everywhere at the same time, and time trends can be controlled for. The authors typically find higher average scores and lower score inequality in comprehensive settings.

Unfortunately enough, comprehensive school reforms tend to include other changes than the postponement of tracking alone, and it is hard to differentiate between their respective effects. For example, we should not be surprised to find that average achievement increases when the lower track is effectively being integrated into the higher one, and the quantity of education is increased for the lower track students.

An interesting paper in this respect is that of Ofer Malamud and Christian Pop-Eleches (2007), which looks at a Romanian comprehensive reform in which the curriculum was changed, but peer group composition was left unaltered. The authors find that although children from disadvantaged backgrounds were more likely to complete the academic track after the reform, they were not more likely to complete university education. Similarly, the Dutch parliamentary report *Parlementair Onderzoek Onderwijsvernieuwin-*

gen (2008) documents how a reform which aimed to implement a common curriculum across a tracked school system, failed. Differences between schools persisted, and the post-reform achievement difference between the tracks was just as large as before.

A second identification strategy is to use regional variance in tracking. This is done when the data does not allow for an intertemporal dimension to the analysis, either because there is no temporal variation in tracking policies, or because only cross-sectional data is available.

One important source of cross-sectional analyzes is the British comprehensive school reform that started in the 1960s. The reform was introduced in different regions at different times, and authors have used a cross-sectional snapshot to show that differences between students increase more in tracked regions (see e.g. Kerckhoff, 1986). From the NCDS data set, test score data is available for the 1958 birth cohort at ages 7, 11 and 16.

International student achievement tests have allowed for cross-country comparisons as well. Eric Hanushek and Ludger Woessmann (2006) use international test data collected in PISA, PIRLS and TIMSS since 1995 to estimate the effect of tracking on the distribution of outcomes. They find lower average achievement and higher score variation in early tracking countries.

Cross-sectional identification strategies are prone to suffer from some kind of omitted variable bias. The educational production function is commonly assumed to be of the following form

$$y_{2i} = \alpha_2 + X_i\beta_2 + U_iv_2 + \gamma_2T_i + \varepsilon_{2i}, \quad (1)$$

where y_2 is late student achievement, typically measured near the end of compulsory education, T a dummy variable indicating early tracking, X a matrix of other, observed variables determining educational outcomes, and U a matrix of unobserved ones. If some of the unobserved (and thus omitted) variables are correlated with T , the estimate of the effect of tracking $\hat{\gamma}_2$ will be biased.

In cross-sectional papers, the omitted variable problem is commonly circum-

vented by using a so-called *value-added* specification. The thought behind it is that the effect of the omitted variables is present in achievement at an earlier age as well. If we have data on early test scores y_1 , we can thus add it to the specification as a control

$$y_{2i} = \alpha_2 + X_i\beta_2 + \gamma_2T_i + \delta_2y_{1i} + \varepsilon_{2i}. \quad (2)$$

In doing so, we implicitly assume that early achievement is determined by its own educational production function

$$y_{1i} = \alpha_1 + X_i\beta_1 + U_iv_1 + \gamma_1T_i + \varepsilon_{1i}, \quad (3)$$

of which the unobserved coefficients v_2 are given by $v_2 = \delta_2v_1$. If v_2 has a different form in reality, the estimate of γ_2 will still be biased. Also, γ_1 must equal zero; tracking must not have an anticipatory effect.

An alternative is to use instrumental variables. In that case, we need an instrument which correlates with T , but not with U . Using the UK data, Galindo-Rueda and Vignoles (2004) instrument tracking policies with political orientation and include a wide range of control variables. They reach the same conclusions as can be found in the earlier literature on the UK reform: tracking increases differences between students, but has a negligible effect on averages.

Allan Manning and Joern-Steffen Pischke (2006) argue that the true effect of tracking cannot reliably be estimated on the basis of the UK comprehensive school reform. The data includes test scores for three different ages, and the authors use this to specify a value-added model with which they estimate an *early* educational production function

$$y_{1i} = \alpha_1 + X_i\beta_1 + \gamma_1T_i + \delta_1y_{0i} + \varepsilon_{1i}. \quad (4)$$

They find that $\hat{\gamma}_1$ is significantly larger than zero. This can partly be explained away (and controlled for) by measurement error in the earliest test, but a substantial portion of the effect remains. The authors then argue that

since γ_1 should be zero in reality, the model must be misspecified. Value-added models for later age achievement, such as (2), are therefore likely to be misspecified as well. Manning and Pischke include a wide range of variables, and like Galindo-Rueda and Vignoles, they also try using political orientation as an instrument, but to no avail. Since it is unlikely that additional data can be found which would solve the problem, this leads them to the conclusion that the true value of γ_2 cannot be known.

The central assumption of Manning and Pischke is that γ_1 is different from zero. If it is not, the specification test is not valid, and we should not reject the existing literature on the basis of it.

3 Anticipatory effects

The effect of tracking may extend to the years preceding actual tracking. Students know that their track placement will largely determine their further educational career. It should therefore be expected that they work harder before the selection point. Likewise, parents and teachers should be expected to push students harder. They may also pressure government into spending more resources on early education. Finally, it is also possible that the signaling effect of track placement increases individual returns to education, causing students to increase overall effort.

There is indeed some existing empirical evidence to support the idea of anticipatory effects. John Bishop (1998, 2001) finds that high-stakes testing correlates positively with student achievement in a number of data sets. High-stakes testing is related to curriculum tracking because many countries base track placement on some kind of test. Even in the absence of a formal test, the incentives emanating from tracking are likely to be similar to those from high-stakes testing.

Though tracking may increase test results, it is not strictly necessary for underlying average achievement to increase as well. Students and teachers may substitute effort in nontested subjects and proficiencies for effort in

tested ones, crowding out nontested subjects. Winters et al. (2008) find evidence that high-stakes testing in math and reading in fact does not crowd out science achievement in Florida schools. They suggest that this may be the case because positive spillover effects from the tested subjects compensate for the crowding-out of nontested ones.

There is also more direct evidence of anticipatory effects of tracking. Fernando Galindo-Rueda and Anna Vignoles (2004) test the relationship between the introduction year of comprehensive schooling and the achievement gain between age 7 and 11 for the 1958 UK birth cohort. They find that the achievement gain was significantly higher for students whose secondary schools turned comprehensive only after they had entered it. In other words: students who had expected an educational career in a tracked secondary school system performed better before entering it.

It is hard to draw firm conclusions from these data only, particularly because the effect Galindo-Rueda and Vignoles observe can have been caused by selection bias. If their result is caused by anticipatory effects of tracking however, we should expect it to be visible in other data as well.

To investigate the matter further, I look at international test score data on fourth graders collected in the PIRLS 2001 and 2006 and in the TIMSS 1995 and 2003 surveys (IEA, 1995, 2001, 2003, 2006). PIRLS is an internationally comparable early age reading literacy survey. TIMSS surveys mathematics and science literacy at three different grades, of which I use the earliest. Both surveys aim to test a representative sample of the population of fourth graders in the participating countries.

Where available, I use the the tracking data of Hanushek and Woessmann (2006). I do this to keep these results as comparable as possible to their paper. I use a larger sample of countries than Hanushek and Woessmann, and I therefore supplement their measure with my own assessment for missing countries. For this I use the Eurybase database (Eurydice 2008), as well as other sources.

I keep as close to Hanushek and Woessmann's definition of tracking as 'age of first tracking' as I can. Just like Hanushek and Woessmann, I use a dummy

variable for the analysis, where a country is considered to be tracking when it does so at an age of 14 or below for the countries in TIMSS, and at 15 or below in PISA/PIRLS. Using a dummy variable is somewhat arbitrary, but not more so than the alternative: to assume that anticipatory effects would be linear in years before the start of tracking.

I pool the data from the different surveys and tests and add dummies to allow for different average scores between tests in the sample. I also discard a number of countries from sample because of ambivalent information on tracking policies.¹ This leaves 52 countries and regions for the estimates.² The sample for which all control variables are available is somewhat smaller at 43.

To take into account the clustered nature of the data, I estimate the effect of tracking using a linear mixed-effect model. The results can be seen from Table 1. Looking at the first specification, the estimates reveal a pattern similar to the UK results. Countries with early tracking clearly have higher score means, with the mean difference as large as 44% of a standard deviation in international student test scores. This finding is significant at the 5% level.

More developed countries may have a different taste for tracking. To control for this, I include real per capita GDP from the Penn World Table (2006). I also include educational expenditures as a percentage of GDP. Countries which value education highly should spend more on it, and should have higher test scores as well. These data are taken from the World Bank EdStat database (2008).

The estimates from this specification can be seen from column (2). Estimated anticipatory effects are now somewhat smaller at 37% of a standard deviation. Unfortunately, the additional variables are not available for all countries.

¹Belize, Kuwait, Qatar, South Africa, Trinidad and Tobago.

²Argentina, Armenia, Australia, Austria, Bulgaria, Canada, Chinese Taipei, Cyprus, the Czech Republic, Denmark, England, Flanders, France, Georgia, Germany, Greece, Hong Kong, Hungary, Iceland, Indonesia, Iran, Ireland, Israel, Italy, Japan, Latvia, Lithuania, Luxembourg, Macedonia, Moldova, Morocco, the Netherlands, New Zealand, Norway, the Philippines, Poland, Portugal, Romania, Russia, Scotland, Singapore, the Slovak Republic, Slovenia, South Korea, Spain, Sweden, Thailand, Tunisia, Turkey, the United States, Yemen, Wallonia.

I rerun the first specification with the more limited sample of countries. The results can be seen from column (3): the full sample has a smaller apparent effect of tracking, but not very much so. I include both GDP and expenditures as controls in the remainder of the analysis.

Children start formal schooling at different ages. The entry age may be related to early tracking, and this could then explain the correlation between tracking and early test scores. Even so, tracking itself may also cause an earlier entry age as a kind of anticipatory effect. I take data on school starting ages from EdStat to see if there is any relationship between tracking and the entry age. As can be seen from column (4), the inclusion of entry age hardly changes the tracking estimate from column (2). The estimate of the effect of the entry age is not significantly different from zero either. I thus find no evidence that the entry age is relevant in the context of tracking.

I try to estimate whether anticipatory effects differ for children with different parental backgrounds. For this, I use a dummy variable which indicates whether the student has less than one case of books at home. Books at home is probably a better measure of parental background than parental education or occupation – the data are derived from a student questionnaire, and young children are often unaware of the education and exact occupation of their parents.

Results can be seen from column (5). Students with less than one case of books at home score lower on average. Even so, the interaction term with the tracking variable is close to zero.³ I thus cannot find evidence for differing anticipatory effects for children with different backgrounds. Additionally, the inclusion of the number of books at home does not change the estimates for the tracking variable very much.

For the sake of completeness, I also add a specification with all variables in column (6). Cross-country differences in the number of books at home do perhaps not explain much of score differences, and indeed, the estimates are very much the same as in specification (2).

³Note that because the interaction includes both a country-level and an individual-level variable, R cannot reliably estimate the standard error.

Another way of analyzing differential anticipatory effects for different kinds of students is to look at quantile effects. I therefore estimate anticipatory effects on different quantiles by means of quantile regression (see e.g. Koenker, 2005). The resulting estimate of quantile effects can be seen from the solid line in Figure 1. I have estimated standard errors with the help of a clustered bootstrap. 10% confidence intervals based on these have been rendered as a shaded area into the figure.

The estimated effect is significantly positive for all quantiles. This suggests that tracking increases early achievement for everyone. The effect seems to be larger for lower quantiles, suggesting that tracking decreases early differences between students. However, the latter finding is not statistically significant, as witnessed by the wide confidence intervals.

Quantile regressions in effect compare the joint student populations of early and late tracking countries. It could be argued that the slope of the solid line represents differences in between-country as well as within-country variation. It may simply be so that late tracking countries have the same within-country variance in achievement, but that there are larger differences between the means of their achievement distributions.

I change the two groups' country distributions to have the same intra-group mean, and rerun the quantile regressions on the resulting distributions. The result can be seen as the dotted line in Figure 1. It is less steep than the original line, but not very much so. This illustrates that the (solid-line) quantile effects are mainly caused by differences between the shape of individual countries' achievement distributions rather than by differences in group level variation.

We have seen from the above that there is a clear link between tracking policies and early test scores. What could be causing this? A first thought would be that the estimates are classical false positives. Their statistical significance shows that this is somewhat unlikely. Also, the occurrence of false positives in international data should be completely independent of such occurrences in the UK analysis of Galindo-Rueda and Vignoles.

A second possibility is that tracking policies are linked to early test scores

through a third variable. It is however not easy to come up with such a variable. In the UK data, a large number of control variables is already included. In the cross-country data, it is hard to think of many mechanisms that link tracking to higher early test scores, other than a direct causal effect.

The most important omitted variable in any educational production function, innate ability, can be assumed to be approximately constant across countries. A second candidate would be countries' levels of economic development, but I already control for per capita GDP and educational expenditures. Other possible controls, the school-starting age and parental background do not change the tracking estimate very much.

Of course, countries with early tracking could be structurally different in other aspects of their educational systems. They may for example have an exogenous preference for competitive, achievement-oriented primary schools, driving both higher early test scores and tracking policies. In my opinion, this is not a methodological problem, but rather changes the interpretation of the estimates. Tracking becomes a proxy for all related policies, and the estimate reflects the effect of such policies on early test scores.

All in all, I feel that anticipatory effects are perhaps the most reasonable explanation for my empirical results. Even so, I hesitate to conclude that a change in tracking policies would have an immediate effect on early test scores, as this would depend on the exact transmission mechanism between tracking and scores.

4 Robustness problems

All cross-country studies on curriculum tracking have robustness problems, and this one is no exception. Waldinger (2006) finds the Hanushek and Woessmann results to be sensitive to variations in the tracking measure, and I confirm this. The lack of robustness is the worse because, as Brunello and Checchi (2007, pp. 800-801) note, the tracking variable is likely to suffer from measurement error as internationally comparable data on the structure of

educational systems is rather scarce. While there are other ways to measure the net effect of tracking, to my knowledge, there exists no set of panel data on early achievement that spans over a tracking reform. Thus this shortcoming seems unavoidable when considering anticipatory effects.

The differences between the tracking measures are not only caused by imprecise information, as the term measurement error might suggest. Authors must also make a number of seemingly arbitrary choices within their broader tracking definition. For example, in the Netherlands and Flanders, students are supposed to follow a common curriculum for up to three years after they are split up into separate schools. Should they be considered tracked from the moment they are split up, or from the moment their curricula formally diverge? What to do with countries where a significant proportion of students drops out of (comprehensive) school? One can also wonder if any of the countries with a formally comprehensive system past the age of 16 are tracking after that age in reality.

The use of a tracking dummy instead of a continuous variable relieves us from the need to make explicit choices in some of the above examples, but some differences remain after the respective measures have been replaced by their dummy equivalents.

To make matters worse, not all authors use the same tracking definition to start with. For example, while Hanushek and Woessmann put the threshold for their dummy just before the age of measurement in the late achievement test (after 14 for TIMSS and after 15 for PISA/PIRLS), Waldinger uses a tracking dummy with a threshold as early as after grade 5, or about age 11. Ammermueller (2005) does not even look at the age tracking starts, but takes the number of tracks at a certain age as a measure of tracking instead.

To illustrate the problem, I construct a reasonable alternative tracking measure parallel to that of Hanushek and Woessmann (2006). The result can be seen from the second column of Table 2. The data are sorted by the underlying continuous tracking variable of Hanushek and Woessmann in the first column. I have also added the tracking grades as reported by Waldinger (2006) and Bedard and Cho (2007) in the third and fourth columns.

The differences between the Hanushek and Woessmann and the alternative tracking measure may seem small. They are however large enough to have sizable effects on tracking estimates. Table 3 shows estimates similar to those of Table 1, but with the original tracking variable replaced with the alternative one. Even though the estimated anticipatory effect of tracking is still positive, it is much smaller and no longer statistically significant.

I also replicate the analysis of Hanushek and Woessmann with the alternative tracking variable. The result can be seen from Tables 4 through 7. The original estimates of the effect on score inequality show a clear pattern, with estimates significantly higher than zero in four out of eight cases. With the alternative tracking measure, only one significant estimate is left.

Changes between the specifications are smaller when considering the effect on mean scores. Three estimates are significantly negative and one is significantly positive in the original specification, while two estimates are significantly negative and one is significantly positive in the alternative specification.

5 Conclusions

This paper investigates possible anticipatory effects of curriculum tracking. Tracking gives students an incentive to work harder in advance because they know that their track placement will largely determine their further educational and professional career.

Anticipatory effects can be estimated empirically. Early test scores are between 0.36 and 0.49 standard deviations higher in early tracking countries. This result is significant at the 5% level, and is robust to the inclusion of control variables like per capita GDP, educational expenditures, the school starting age and the number of books at home.

I look at differential anticipatory effects across achievement quantiles using quantile regression. Anticipatory effects seem to be higher for lower achievers, even though the uncertainty of the estimate is large. I find no evidence

for sizable differential anticipatory effects across groups of different parental backgrounds.

The evidence for the existence of anticipatory effects presented in this paper has implications for previous estimates of the net effect of curriculum tracking. So-called *value-added* models are based on the assumption that no such effects exist. This paper illustrates that it would be prudent to avoid such models when estimating the net effect of tracking. If we suspect that estimates based on late test scores only are susceptible to selection bias, we should instead try to ameliorate the problem by using instrumental variables or panel data.

The specification test of Manning and Pischke is explicitly based on the a priori assumption that γ_1 cannot be zero, i.e. that anticipatory effects cannot exist. I argue that the existence of anticipatory effects is plausible, especially in the light of the evidence presented in this paper. We should therefore not reject the UK literature on the basis of their test.

In the cross-country analysis of Hanushek and Woessmann (2006), early test scores are used to control for all variables in X except per capita GDP and educational expenditures. When omitting early scores, their estimates are no longer significant. If anticipatory effects exist, the latter specification should however be preferred, and we should not reject their null hypothesis that tracking has no effect on (late) mean test scores.

I confirm the earlier finding that at cross-country estimates of the effect of tracking are sensitive to its exact definition. This paper is no exception. Lack of robustness is another reason to mistrust cross-country analyzes, and to put more faith in panel data.

References

- Andreas Ammermueller. Educational opportunities and the role of institutions. ZEW discussion paper no. 05-44, 2005.
- Andreas Ammermueller and Joern-Steffen Pischke. Peer effects in European

- primary schools: evidence from PIRLS. ZEW discussion paper no. 06-027, 2006.
- Kelly Bedard and Insook Cho. The gender test score gap across OECD countries. Working paper, September 2007.
- John Bishop. The effect of curriculum-based external exit systems on student achievement. *Journal of Economic Education*, 29(2):171–182, 1998.
- John Bishop. The effect of national standards and curriculum-based exams on achievement. *The American Economic Review*, 87(2):260–264, 2001.
- Giorgio Brunello and Daniele Checchi. Does school tracking affect equality of opportunity? New international evidence. *Economic Policy*, 52:781–861, 2007.
- Eurydice information network on education in Europe. Eurydice database on education systems in Europe. <http://www.eurydice.org>, 2008.
- Fernando Galindo-Rueda and Anna Vignoles. The heterogeneous effect of selection in secondary schools: understanding the changing role of ability. IZA discussion paper no. 1245, August 2004.
- Eric Hanushek and Ludger Woessmann. Does educational tracking affect performance and inequality? Differences-in-differences evidence across countries. *The Economic Journal*, 116:C63–C76, 2006.
- Caroline Hoxby. Peer effects in the classroom: learning from gender and race variation. NBER working paper no. 7867, August 2000.
- International Association for the Evaluation of Educational Achievement IEA. Trends in International Mathematics and Science Study TIMSS. 1995.
- International Association for the Evaluation of Educational Achievement IEA. Progress in International Reading Literacy Study PIRLS. 2001.

- International Association for the Evaluation of Educational Achievement
IEA. Trends in International Mathematics and Science Study TIMSS.
2003.
- International Association for the Evaluation of Educational Achievement
IEA. Progress in International Reading Literacy Study PIRLS. 2006.
- Alan Kerckhoff. Effects of ability grouping in British secondary schools.
American Sociological Review, 51(6):842–858, 1986.
- Roger Koenker. *Quantile Regression*. Cambridge University Press, 2005.
- Ofer Malamud and Christian Pop-Eleches. The effect of postponing tracking
on access to higher education: evidence from a regression-discontinuity
design. Working paper., Oktober 2007.
- Alan Manning and Joern-Steffen Pischke. Comprehensive versus selective
schooling in England and Wales: what do we know? NBER working
paper no. 12176, April 2006.
- Costas Meghir and Maarten Palme. Educational reform, ability and family
background. *American Economic Review*, 95(1):414–424, 2005.
- Commissie Parlementair Onderzoek Onderwijsvernieuwingen. *Parlementair
Onderzoek Onderwijsvernieuwingen*. Sdu Uitgevers, 2008.
- Tuomas Pekkarinen. Gender differences in educational attainment: evidence
on the role of the tracking age from a Finnish quasi-experiment. IZA
discussion paper no. 1897, December 2005.
- Tuomas Pekkarinen, Roope Uusitalo, and Sari Pekkala. School tracking and
development of cognitive skills. NBER working paper 7867, 2007.
- Penn World Table PWT. Penn world table version 6.2. Alan Heston, Robert
Summers and Bettina Aten; Center for International Comparisons of Pro-
duction, Income and Prices at the University of Pennsylvania, September
2006.

Fabian Waldinger. Does tracking affect the importance of family background on students' test score. Unpublished manuscript, LSE, January 2006.

Marcus Winters, Jay Greene, and Julie Trivitt. The impact of high-stakes testing on student proficiency in low-stakes subjects. Manhattan institute for policy research, Civic report no. 54., July 2008.

World Bank. EdStat Education Statistics. 2008.

Figures and Tables

Dependent variable: early test scores						
	(1)	(2)	(3)	(4)	(5)	(6)
intercept	466.36*** <i>10.62</i>	382.87*** <i>29.76</i>	463.72*** <i>12.18</i>	342.04*** <i>98.58</i>	410.04*** <i>29.12</i>	366.11*** <i>96.41</i>
early tracking	44.21** <i>18.56</i>	36.54** <i>17.57</i>	49.47** <i>19.97</i>	35.73* <i>17.85</i>	37.17** <i>17.19</i>	36.28** <i>17.46</i>
real per capita GDP		4.49*** <i>1.05</i>		4.72*** <i>1.18</i>	4.2*** <i>1.03</i>	4.44*** <i>1.16</i>
expenditures		5 <i>5.32</i>		4.42 <i>5.54</i>	4.09 <i>5.21</i>	3.47 <i>5.42</i>
entry age				6.54 <i>15.05</i>		7.04 <i>14.72</i>
<1 case of books					-30.09*** <i>0.28</i>	-30.09*** <i>0.28</i>
tracking*books					-3.01 <i>NA</i>	-3.01 <i>NA</i>
observations	864155	654729	654729	654729	654729	654729
countries	52	43	43	43	43	43

Table 1: Tracking and early test scores; pooled regression including dummy variables for the different surveys (not shown). Standard errors in italics, p-values: *** 1% ** 5% * 10%. The international student test score standard deviation is 100. Data: IEA 1995, 2001, 2003, 2006; Penn World Table, 2006; Hanushek and Woessmann (2006); Eurydice, 2008.

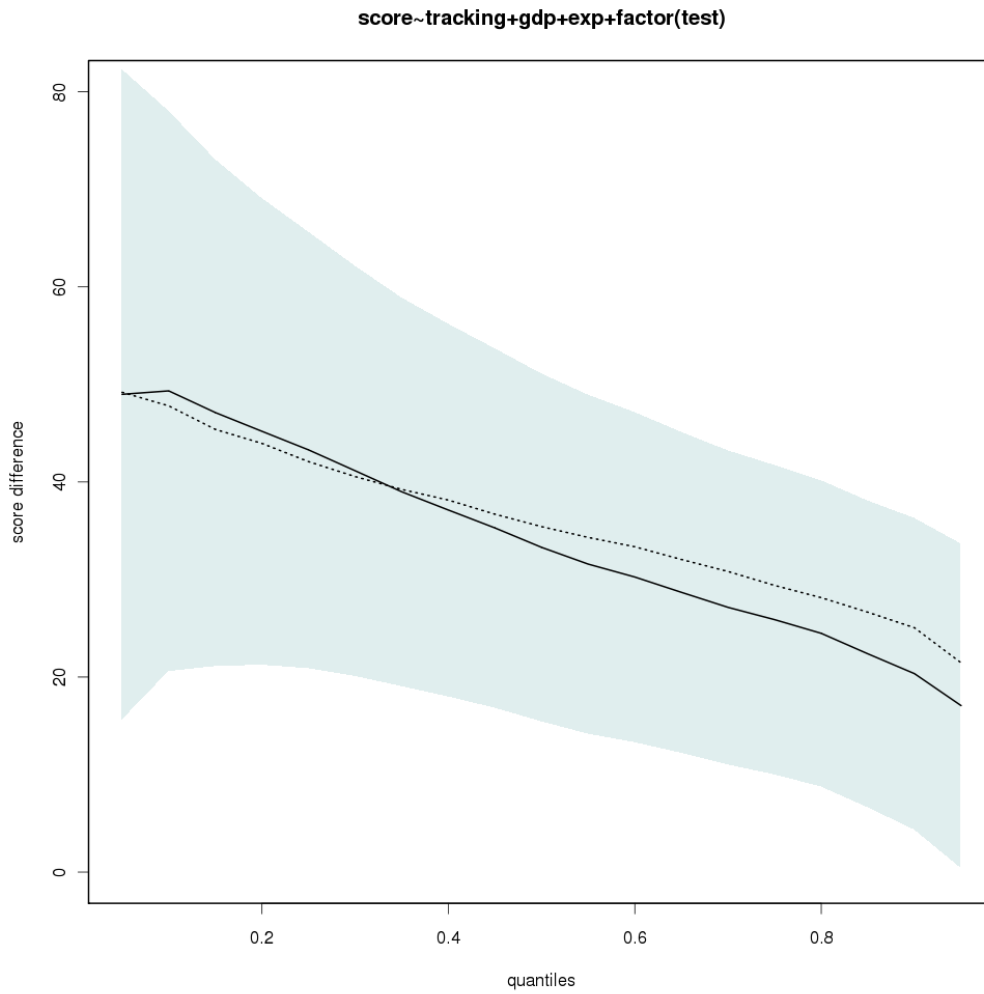


Figure 1: Apparent anticipatory effects of curriculum tracking as estimated by quantile regression. The shaded area depicts the 90% confidence interval. The dotted line includes country-level fixed effects. The control variables from specification (2) have been included. Data: IEA, 1995, 2001, 2003, 2006; Penn World Table, 2006; Eurydice, 2008.

	Hanushek and Woessmann	Alternative assessment	Waldinger	Bedard and Cho
country	age	age	grade	grade
Hungary	10	10	4	4
Austria	10	10	4	4
Germany	10	10	4	4
Slovakia	10	11	4	4
Czech Rep.	11	11	5	5
Netherlands	12	12	6	8
Ireland	12	12	6	11
Israel	12	12		
Philippines	12	13		
Singapore	12	13		
Wallonia	12	14		
Flanders	14	12	6	8
Italy	14	14	8	8
Lithuania	14	14		
South Korea	14	15	9	9
Bulgaria	14	15		
Switzerland	15	15	6	9
Portugal	15	15	9	6
Greece	15	15	9	9
France	15	15	9	9
Romania	15	15		
Cyprus	15	15		
Taiwan	15	15		
Macedonia	15	15		
Russia	15	15		
Slovenia	15	15		
Armenia	15	16		
Indonesia	15	16		
Poland	15	16		
Denmark	16	16	9	9
Norway	16	16	9	10
Iceland	16	16	9	10
Latvia	16	16		
Sweden	16	19	9	9
Moldova	17	15		
Argentina	18	13		
Turkey	18	14	8	
Iran	18	14		
Thailand	18	15		
New Zealand	18	16	6	11
Canada	18	16	8	11
Australia	18	16	9	11
England	18	16	12	11
US	18	16	12	11
Scotland	18	16	12	11
Jordan	18	16		
Malaysia	18	16		
Morocco	18	16		
Japan	18	19	9	9
Hong Kong	19	16		

Table 2: Equivalent tracking measures by Hanushek and Woessmann, 2006, Koerselman, Waldinger, 2006 and Bedard and Cho, 2007. Note that actual analysis is based on dummy variables.

Dependent variable: early test scores						
	(1)	(2)	(3)	(4)	(5)	(6)
intercept	476.81*** <i>11.32</i>	398.62*** <i>32.46</i>	476.28*** <i>13.23</i>	335.54*** <i>104.93</i>	424.84*** <i>31.69</i>	359.65*** <i>102.39</i>
early tracking	11.56 <i>19.24</i>	6.02 <i>18.26</i>	14.79 <i>21.04</i>	6.49 <i>18.42</i>	8.83 <i>17.83</i>	9.32 <i>17.98</i>
real per capita GDP		5*** <i>1.08</i>		5.32*** <i>1.2</i>	4.67*** <i>1.06</i>	5*** <i>1.17</i>
expenditures		2.61 <i>5.69</i>		1.89 <i>5.85</i>	1.84 <i>5.56</i>	1.09 <i>5.71</i>
entry age				9.95 <i>15.72</i>		10.28 <i>15.34</i>
<1 case of books					-30.01*** <i>0.27</i>	-30.01*** <i>0.27</i>
tracking*books					-3.31 <i>NA</i>	-3.31 <i>NA</i>
observations	864155	654729	654729	654729	654729	654729
countries	52	43	43	43	43	43

Table 3: Tracking estimates for an alternative tracking measure, specifications as in Table 1 Data: IEA 1995, 2001, 2003, 2006; Penn World Table, 2006; Eurydice, 2008.

	PISA 00 PIRLS reading	PISA 03 PIRLS reading	TIMSS 03 TIMSS 03 math	TIMSS 03 TIMSS 03 science
early tracking	-0.018 (0.077)	0.248** (0.110)	0.013 (0.054)	0.105 (0.073)
early inequality	0.255* (0.139)	0.594*** (0.129)	-0.014 (0.248)	0.252 (0.176)
	TIMSS 95 TIMSS 95 math	TIMSS 95 TIMSS 95 science	TIMSS 99 TIMSS 95 math	TIMSS 99 TIMSS 95 science
early tracking	0.147* (0.076)	0.197** (0.084)	0.005 (0.074)	0.208* (0.107)
early inequality	0.476 (0.306)	0.843*** (0.224)	0.099 (0.146)	0.785*** (0.135)

Table 4: Tracking and late score inequality, source: Hanushek and Woessmann, 2006. Dependent variable: secondary school inequality as measured by the weighted standard deviation of test scores. Early age inequality is included as a control. Standard errors within parentheses. Significance levels: *** 1% ** 5% * 10%. Note that the authors have divided the estimates by 100.

	PISA 00 PIRLS reading	PISA 03 PIRLS reading	TIMSS 03 TIMSS 03 math	TIMSS 03 TIMSS 03 science
early tracking	-1.88 (2.85)	0.21 (2.30)	-1.50 (3.15)	2.68 (3.17)
early inequality	0.39*** (0.11)	0.51*** (0.12)	0.11 (0.13)	0.28*** (0.07)
	TIMSS 95 TIMSS 95 math	TIMSS 95 TIMSS 95 science	TIMSS 99 TIMSS 95 math	TIMSS 99 TIMSS 95 science
early tracking	-2.10 (4.48)	1.91 (3.80)	-2.12 (3.68)	7.58* (4.09)
early inequality	0.47* (0.27)	0.61*** (0.19)	0.02 (0.19)	0.52** (0.19)

Table 5: Tracking and late score inequality, source: Koerselman. Dependent variable: secondary school inequality as measured by the weighted standard deviation of test scores. Early age inequality is included as a control. Standard errors within parentheses. Significance levels: *** 1% ** 5% * 10%. Data: IEA 1995, 2001, 2003, 2006.

	PISA 00 PIRLS reading	PISA 03 PIRLS reading	TIMSS 03 TIMSS 03 math	TIMSS 03 TIMSS 03 science
early tracking	-0.951*** (0.287)	-1.053*** (0.343)	0.021 (0.157)	-0.013 (0.161)
early means	0.643*** (0.130)	0.676*** (0.139)	0.928*** (0.085)	0.929*** (0.075)
	TIMSS 95 TIMSS 95 math	TIMSS 95 TIMSS 95 science	TIMSS 99 TIMSS 95 math	TIMSS 99 TIMSS 95 science
early tracking	-0.062 (0.135)	0.597** (0.222)	-0.410* (0.219)	0.234 (0.370)
early means	0.965*** (0.063)	0.738*** (0.097)	1.045*** (0.088)	0.828*** (0.124)

Table 6: Tracking and late score means, source: Hanushek and Woessmann, 2006. Dependent variable: secondary school weighted score means. Early age means are included as a control. Standard errors within parentheses. Significance levels: *** 1% ** 5% * 10%. Note that the authors have divided the estimates by 100.

	PISA 00 PIRLS reading	PISA 03 PIRLS reading	TIMSS 03 TIMSS 03 math	TIMSS 03 TIMSS 03 science
early tracking	-31.34** (15.38)	-26.14** (8.84)	0.66 (10.16)	3.77 (9.60)
early means	0.63*** (0.21)	0.41** (0.17)	0.84*** (0.07)	0.71*** (0.07)
	TIMSS 95 TIMSS 95 math	TIMSS 95 TIMSS 95 science	TIMSS 99 TIMSS 95 math	TIMSS 99 TIMSS 95 science
early tracking	-0.14 (7.57)	27.79*** (8.08)	-12.57 (12.60)	16.37 (13.12)
early means	1.01*** (0.08)	0.72*** (0.10)	0.93*** (0.12)	0.68*** (0.14)

Table 7: Tracking and late score means, source: Koerselman. Dependent variable: secondary school weighted score means. Early age means are included as a control. Standard errors within parentheses. Significance levels: *** 1% ** 5% * 10%. Data: IEA 1995, 2001, 2003, 2006.

Aboa Centre for Economics (ACE) was founded in 1998 by the departments of economics at the Turku School of Economics, Åbo Akademi University and University of Turku. The aim of the Centre is to coordinate research and education related to economics in the three universities.

Contact information: Aboa Centre for Economics, Turku School of Economics, Rehtorinpellonkatu 3, 20500 Turku, Finland.

Aboa Centre for Economics (ACE) on Turun kolmen yliopiston vuonna 1998 perustama yhteistyöelin. Sen osapuolet ovat Turun kauppakorkeakoulun kansantaloustieteen oppiaine, Åbo Akademin nationalekonomi-oppiaine ja Turun yliopiston taloustieteen laitos. ACEn toiminta-ajatuksena on koordinoida kansantaloustieteen tutkimusta ja opetusta Turun kolmessa yliopistossa.

Yhteystiedot: Aboa Centre for Economics, Kansantaloustiede, Turun kauppakorkeakoulu, 20500 Turku.

www.ace-economics.fi

ISSN 1796-3133