Government
Institute for
Economic Research

Working Papers 2

School tracking and development of cognitive skills

Tuomas Pekkarinen

Roope Uusitalo

Sari Kerr

Working Papers 2 March 2009

VATT WORKING PAPERS

2

School tracking and development of cognitive skills

Tuomas Pekkarinen
Roope Uusitalo
Sari Kerr

Valtion taloudellinen tutkimuskeskus
Government Institute for Economic Research
Helsinki 2009

Tuomas Pekkarinen (Helsinki School of Economics, VATT, IZA and IFAU) (e-mail: tuomas.pekkarinen@hse.fi)

Roope Uusitalo (Government Institute for Economic Research (VATT), IZA and IFAU) (e-mail: roope.uusitalo@vatt.fi)

Sari Kerr (Charles River Associates and Boston University) (e-mail: skerr@srai.com)

ISBN 978-951-561-841-2 (nid.)
ISBN 978-951-561-842-9 (PDF)

ISSN 1798-0283 (nid.)
ISSN 1798-0291 (PDF)

Valtion taloudellinen tutkimuskeskus
Government Institute for Economic Research
Arkadiankatu 7, 00100 Helsinki, Finland

Oy Nord Print Ab
Helsinki, March 2009

School tracking and development of cognitive skills

Government Institute for Economic Research
VATT Working Papers 2/2009

Tuomas Pekkarinen – Roope Uusitalo – Sari Kerr

Abstract

The Finnish comprehensive school reform replaced the old two-track school system with a uniform nine-year comprehensive school and significantly reduced the degree of heterogeneity in the Finnish primary and secondary education. We estimate the effect of this reform on the test scores in the Finnish Army Basic Skills test. The identification strategy relies on a differences-in-differences strategy and exploits the fact that the reform was implemented gradually across the country during a six-year period between 1972 and 1977. We find that the reform had a small positive effect on the verbal test scores but no effect on the mean performance in the arithmetic or logical reasoning tests. Still in all tests the reform improved the scores of students from families where parents had only basic education.

Key Words: education, school system, tracking, comprehensive school, test scores

JEL Codes: H52, I21

Tiivistelmä

1970-luvulla toteutettu peruskoulu-uudistus korvasi vanhan rinnakkaiskoulu-järjestelmän yhdeksänvuotisella peruskoululla. Vastaavia uudistuksia on toteutettu aiemmin Ruotsissa ja Norjassa ja Isossa-Britanniassa. Rinnakkaiskoulun ja yhtenäis-koulun eduista ja haitoista keskustellaan edelleen kiivaasti muissa Euroopan maissa.

Tässä tutkimuksessa arvioidaan peruskoulu-uudistuksen vaikutuksia armeijan peruskoe 1:ssä. Tutkimusasetelma perustuu uudistuksen asteittaiseen toteuttamiseen eri kunnissa vuosien 1972 ja 1977 välillä. Tulosten mukaan peruskoulu-uudistuksella oli pieni positiivinen vaikutus verbaalista lahjakkuutta mittaavassa testissä, mutta ei vaikutusta keskimääräisiin tuloksiin matemaattista lahjakkuutta tai loogista päättelykykyä mittaavissa testeissä. Kaikissa testeissä peruskoulu näytti kuitenkin parantaneen testituloksia niitten oppilaitten joukossa, joitten vanhemmilla oli vain perusasteen koulutus.

Asiasanat: peruskoulu, oppimistulokset

JEL-luokittelu: H52, I21

1. Introduction

Publication of international comparisons of student achievement, such as the OECD's Programme for International Student Assessment (PISA), has generated growing interest in the effect of the schooling systems on student outcomes. For reasons not yet fully understood, students in some countries seem to perform substantially better than students in other countries with roughly equal school resources. In addition, there seems to be no clear efficiency-equity trade-off. For example, the Finnish students have consistently been top performers in the recent PISA studies. Still, also the variance in test scores is among the smallest in Finland.

One potential explanation for cross-country differences has to do with the timing and the extent of tracking or ability grouping of students. Comprehensive and selective school systems differ importantly in this respect. In the selective system, tracking into different types of schools occurs early and choices made as early as around age ten largely determine later schooling options. In the comprehensive system, on the other hand, students often stay in the same schools until the end of secondary school.

In this paper, we study the effect of the comprehensive school system on cognitive test scores. Our strategy is to exploit the Finnish comprehensive school reform that gradually transformed the Finnish school system from a selective two-track system to a comprehensive school system that is similar to most European comprehensive school systems. We estimate the effect of this reform on mathematical, verbal, and logical reasoning tests that the Finnish army recruits take during the first weeks of the compulsory military service.

Economic theory suggests that ability grouping may be beneficial for learning since it allows schools to target teaching to more homogeneous groups (Lazear, 2001). This helps both the best and the weakest students. However, if the peer groups have a large effect on student achievement, the students that are placed to lower-ability tracks may also suffer from ability grouping. Since reallocation of students across schools necessarily implies that some students experience an increase in the average peer quality while other students experience an equal decrease, the overall benefits of tracking also depend on whether the peer-effects are linear or not, i.e. whether bad students hamper learning more than good students improve it. A less controversial conclusion is that if the average peer group quality affects learning, the variance in student performance should be higher with ability grouping.

Empirical literature has made a lot of progress on understanding the effects of tracking within schools. Most recent evidence comes from a randomized experiment in Kenya where randomly selected schools implemented tracking and non-tracking policies. Duflo, Dupas, and Kremer (2008) show that tracking within schools seems to benefit all students and they attribute this effect to better targeting of teaching that is made possible by more homogeneous groups. As noted by the authors it is not clear whether the results would be similar in developed countries where student population is less heterogeneous and where additional resources such as remedial education and special education programs may already help teachers to deal with different types of students. In addition, ability grouping is not the only difference between selective and comprehensive school systems. In a selective system, also teacher quality is likely to differ between schools and the curriculum is typically quite different in high ability and low ability tracks. Hence, even a well designed randomized experiment of tracking within schools is unlikely to settle the policy question whether the school system should be selective or comprehensive.

In the developed countries, most existing evidence on the potential benefits of selective vs comprehensive system originates from cross-country comparisons. For example, Hanushek and Wößmann (2006) use data from international student assessments and find that the variance in the test scores is higher in the countries where tracking takes place at an early age. At the same time, early tracking seems to have generally negative effects on mean performance, though the evidence on these mean effects is less consistent. A similar cross-country approach is used by Brunello and Checchi (2007) and Waldinger (2006). Neither of these studies confirms the Hanushek -Wößmann results according to which early tracking increases inequality in achievement

Conflicting results from previous studies reflect the difficulties in analysing the effect of school system based on cross-country data. While these studies try to control for variation due to other factors by including early test scores (Hanushek and Wößmann, 2006; and Waldinger, 2006) or by using time variation in the tracking age (Brunello and Checchi, 2007) it is far from clear that all relevant cross-country differences would be accounted for.

We claim that analysing the changes in test scores when a country switches from a tracked to a comprehensive system is a more promising approach to identify the effects of the school system. Previous attempts to do this include Kerckhoff et al (1996) as well as Galindo-Rueda and Vignoles (2005) who both study the effect of gradual movement from selective school

system into a comprehensive system in England and examine the effects on cognitive ability measured in the National Child Development Survey (NCDS). However, as noted by Manning and Pischke (2006) the areas that first switched to a comprehensive system in England were on average poorer than the areas which retained the tracked system. It is therefore difficult to distinguish the effect of school systems from these regional differences using a single cohort data such as the NCDS. Manning and Pischke also demonstrate that controlling for pre-tracking test scores or using arguably exogenous variation in the timing of the reform as an instrument do not solve the problem due to endogenous timing of the reform.

Compared to these studies the distinct advantage of the Finnish reform is the availability of comparable data from several cohorts, which avoids the need to rely on the cross-sectional variation only. The Finnish comprehensive school reform was implemented gradually region by region between 1972 and 1977. This gradual implementation and the availability of data on several cohorts allow controlling for regional variation and any time trends in the student achievement using a difference-in-differences approach. Furthermore, our data also include information on families, which makes it possible to estimate the effect of the reform based on data on brothers who were placed into different school systems. We estimate the effect of this reform on the cognitive test scores that the Finnish army recruits take when they first enter the mandatory military service of which very few young men are exempted. We find that the reform had a small positive effect on the verbal test scores but no effect on the mean performance in the arithmetic or logical reasoning tests. Still, the reform had a positive effect on test scores in all tests in a group of students from families where parents had only basic education. The reform did not significantly reduce the overall variance of the test scores.

We are not the first to study the effects of comprehensive school reforms. However, none of the earlier papers addresses the effects of the school system on the skill distribution. For example, Meghir and Palme (2005) examine the effects of a Swedish comprehensive school reform on educational attainment and earnings. They find that the Swedish reform increased the earnings and completed schooling of the children of working-class families. In our previous research (Pekkarinen, Uusitalo and Pekkala, 2006) we examine the effects of the Finnish comprehensive school reform on intergenerational income elasticity and find that income mobility increased after the reform. Holmlund (2008) reports similar results based on the Swedish comprehensive school reform. Black, Deveraux, and Salvanes (2005) report that the Norwegian comprehensive school reform increased educational attainment. While all

these papers report important effects on various labor market outcomes, they are silent about the mechanisms behind the results. By measuring the effects on productive skills we may get a step closer to understanding how differences in school systems affect learning.

The rest of the paper is organized as follows. In the following section, we describe the comprehensive school reform in detail. The third section presents the data. Particular emphasis is put on the description of the cognitive tests of the army. In section four, we describe our empirical approach and in section five present the results. The sixth section concludes with some final comments.

2. Comprehensive school reform[1]

2.1 Background

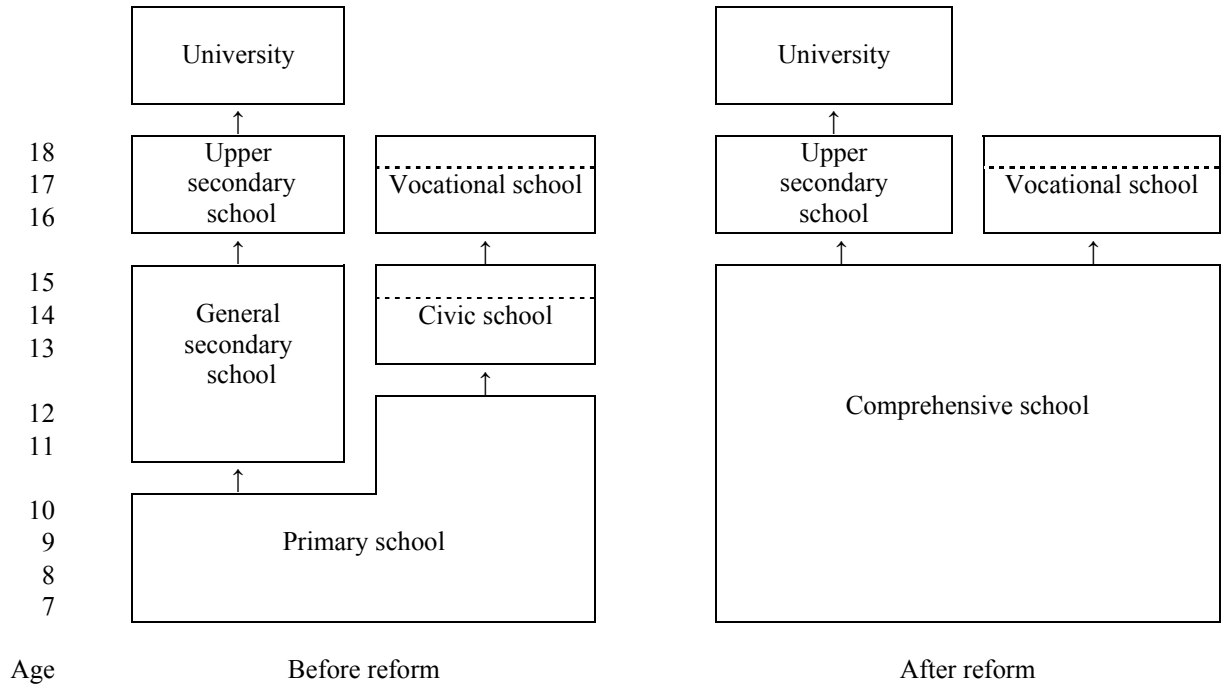
Finland introduced a wide-ranging comprehensive school reform in the 1970's. Similar reforms had already taken place in Sweden in 1950s and in Norway in 1960s (Meghir and Palme, 2005; Aalvik, Salvanes and Vaage, 2003). The Finnish comprehensive school reform abolished the old two-track school system and created a uniform 9-year comprehensive school. The main motivation of the reform was to provide equal educational opportunities to all students, irrespective of place of residence or social background.

In the pre-reform system all students entered primary school ("kansakoulu") at the age of seven. After four years in the primary school, at age 11, the students were faced with the choice of applying to general secondary school ("oppikoulu") or continuing in the primary school. Admissions to the general secondary school were based on an entrance examination, a teacher assessment and primary school grades. Those who were admitted continued their schooling in the junior secondary schools for five years and often went on to the upper secondary school for three additional years. At the end of the upper secondary school the students took the matriculation examination that provided eligibility to university-level studies. Those who were not admitted or who did not apply to the general secondary school continued in primary school for two more years, and spent in total six years in the primary school. By the beginning of 1970s most primary schools had continuation classes or civic

[1] This section draws on our previous paper (Pekkarinen et al 2006)

schools that kept almost the entire age cohort at school up to the 8th (and in many municipalities 9th) grade. This education did not provide eligibility for senior secondary school or for university studies. After civic school most students continued into vocational education or discontinued their schooling. The pre-reform system is described schematically in the left-hand panel of Figure 1.

Figure 1 Finnish school systems before and after the comprehensive school reform



2.2 Content of the comprehensive school reform

The reform introduced a new curriculum and changed the structure of primary and secondary education. The new curriculum increased the academic content of education compared to the old primary school curriculum by increasing the share of mathematics and sciences. In addition, one foreign language became compulsory for all students. Thus, the new comprehensive school curriculum resembled the old general secondary school curriculum and exposed the pupils who, in the absence of the reform would have stayed in the primary school, to a significantly more academic education.

The structure of the post-reform school system is described in the right-hand panel of Figure 1. Previous primary school, civic school and junior secondary school were replaced by a nine-

year comprehensive school. At the same time the upper secondary school was separated from the junior secondary school to form a distinct form of institution. Thus, after the reform, all the pupils followed the same curriculum in the same establishments (comprehensive schools) up to age 16. After this, the students chose between applying to upper secondary school or to vocational schools. Admission to both tracks was based solely on comprehensive school grades.

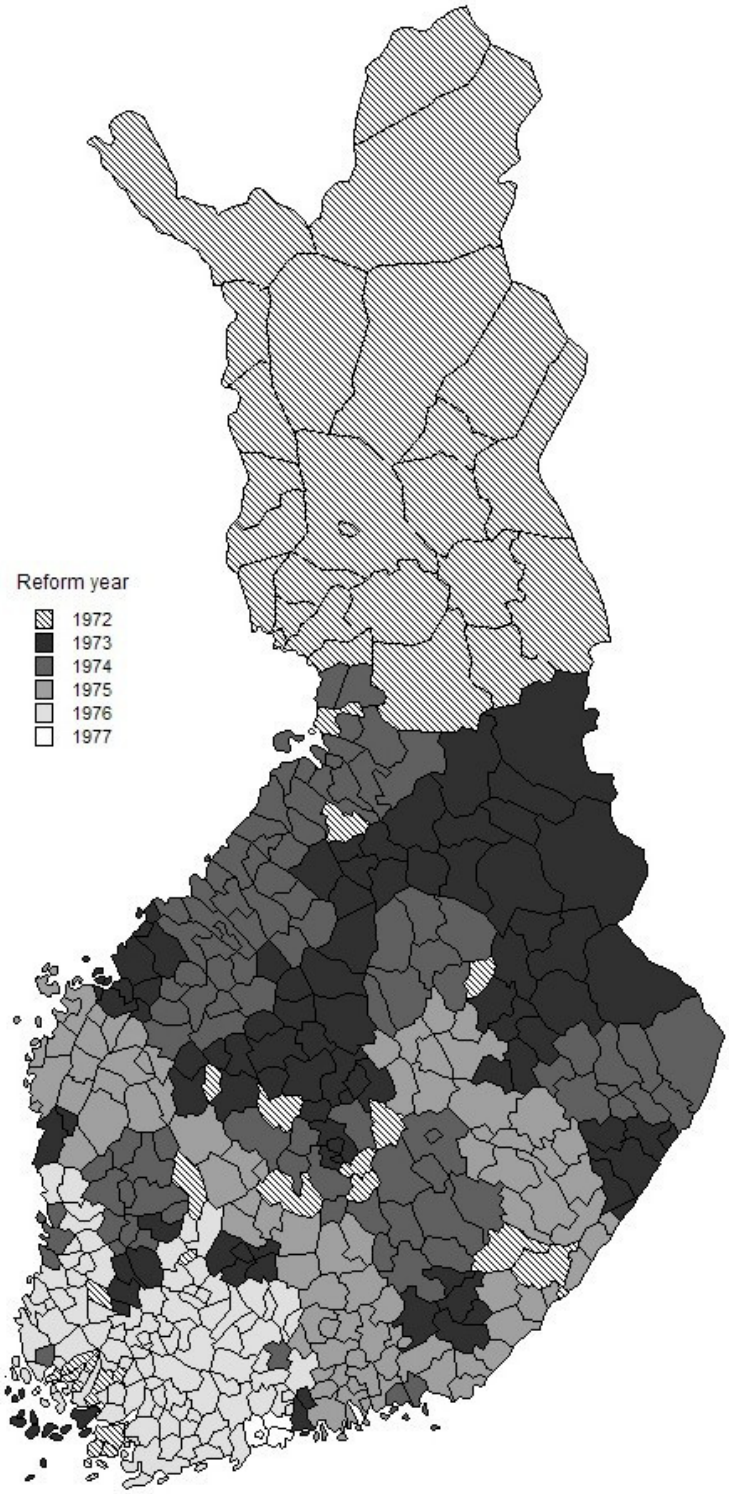
2.3 The implementation of the comprehensive school reform

The implementation of the reform was preceded by a process of planning that lasted for two decades. Government working groups had proposed creating a comprehensive school already in 1948, 1957, 1959, and 1965. The first experimental comprehensive schools started their operation in 1967. Finally, in 1968 the Parliament approved the School Systems Act (467/1968) according to which the two track school system would be gradually replaced with a nine-year comprehensive school. The adoption of the new school system was to take place between 1972 and 1977, and the order in which the municipalities adopted the reform was to be determined by geography starting from the Northern Finland where access to education was most limited. A regional implementation plan divided the country into implementation regions and dictated when each region would adopt the comprehensive school system.

In each region, the five lowest primary school grades were to start in the comprehensive school immediately during the fall term of the year stated in the implementation plan. After this, each incoming cohort of first graders would start their schooling in the comprehensive school. The pupils that were already above the fifth grade in the year that the region entered the reform would complete their schooling according to the pre-reform system. Thus, in each region it took approximately four years to complete the reform.

Figure 2 illustrates how the reform spread through the Finnish municipalities during 1972-1977. The first municipalities that adopted the reform in 1972 were predominantly situated in the northernmost province of Lapland. In 1973 the reform was mostly adopted in the north-eastern regions. From thereon, the reform spread so that it was adopted in 1974 in the northwest, in 1975 in south-east, in 1976 in the south-west, and finally, in 1977 in the capital region of Helsinki.

Figure 2 The implementation of the comprehensive school reform across regions 1972-1977



2.4 The comprehensive school reform as a quasi-experiment

The Finnish comprehensive school reform is in many ways a promising natural experiment for evaluating the effects of early versus late tracking on student outcomes. A particularly useful setup was created by the regional implementation plan that dictated when each municipality moved into comprehensive school system. Using a fixed-effects approach we can control for other simultaneous time trends and regional differences and purge the estimate of school system from these confounding factors.

However, there are some caveats to the approach. First of all, as is clear from Figure 2, some municipalities implemented the reform earlier that the rest of the municipalities in the region. Although the timing of the reform followed the national implementation plan, the choice of first municipalities to implement the reform was probably not entirely random. The comprehensive school reform also faced intensive resistance. Most common arguments against the reform were that abolishing tracking would reduce the quality of education. As a compromise, ability tracking was partially retained within the comprehensive school. Even after the reform the students were divided into ability groups in foreign language and math classes, but studied all other subjects in their regular (not tracked) classes. This ability grouping was eventually abolished in 1985. The socialization of private general secondary schools under municipal ownership was also opposed especially in Helsinki where some of these schools had a distinguished reputation. After an intensive debate, it was agreed that several private schools would be allowed to survive as private alternatives to the comprehensive schools in the Helsinki region even after the reform. Many of these still exist as private senior secondary schools. However, also these private schools had to follow the national curriculum, were funded by the municipalities and could not charge student fees. Hence, the distinction between private and public schools was never very important in Finland.

3. Data

A fundamental problem in assessing the effects of a school reforms on student performance is that students in separate school systems rarely participate in comparable tests. Sometimes it is possible to use nation-wide or international comparisons of student achievement. However,

since most large-scale school reforms took place in 1960s and 1970s when testing was not as widespread as today, it is difficult to find tests implemented to representative and reasonably large samples of students from both pre- and post reform school systems.

In this paper we use the basic skills test results from the Finnish Army. Since military service is mandatory in Finland almost the entire male cohort takes the test. The Army test is given to all new conscripts between the third and the sixth week of their service. The average age at the time when these men are tested is 20, so obviously also other factors than school system may have had an effect on the results. On the other hand, we are probably more interested in long-lasting outcomes of school systems than the immediate effects on test results. The Finnish Army test is also a strong predictor of earnings and occupation later in life, so any effect of school system on the test scores will have important consequences for lifetime earnings.

The Finnish Army Basic Skills test is designed to measure general abilities. The Army uses the test results in selecting conscripts to officer training. The test consists of three subtests: verbal, arithmetic, and logical reasoning. Each subtest includes forty multiple choice questions sorted in increasing order of difficulty. In the verbal reasoning subtest, the subject has to choose synonyms or antonyms of given words, select words that belong to the same category as a given word, exclude words from a group of words, and to identify similar relationships between word pairs. The arithmetic reasoning test asks the subject to complete number series, solve verbally expressed mathematical problems, compute simple arithmetic operations, and to choose similar relationships between pairs of numbers. The logical reasoning test is a standard "culture free" intelligence test based on Raven's progressive matrices and its results should therefore be less affected by pre-test schooling.[2] On the other hand, both the verbal and arithmetic reasoning parts test skills that are primarily taught in school.

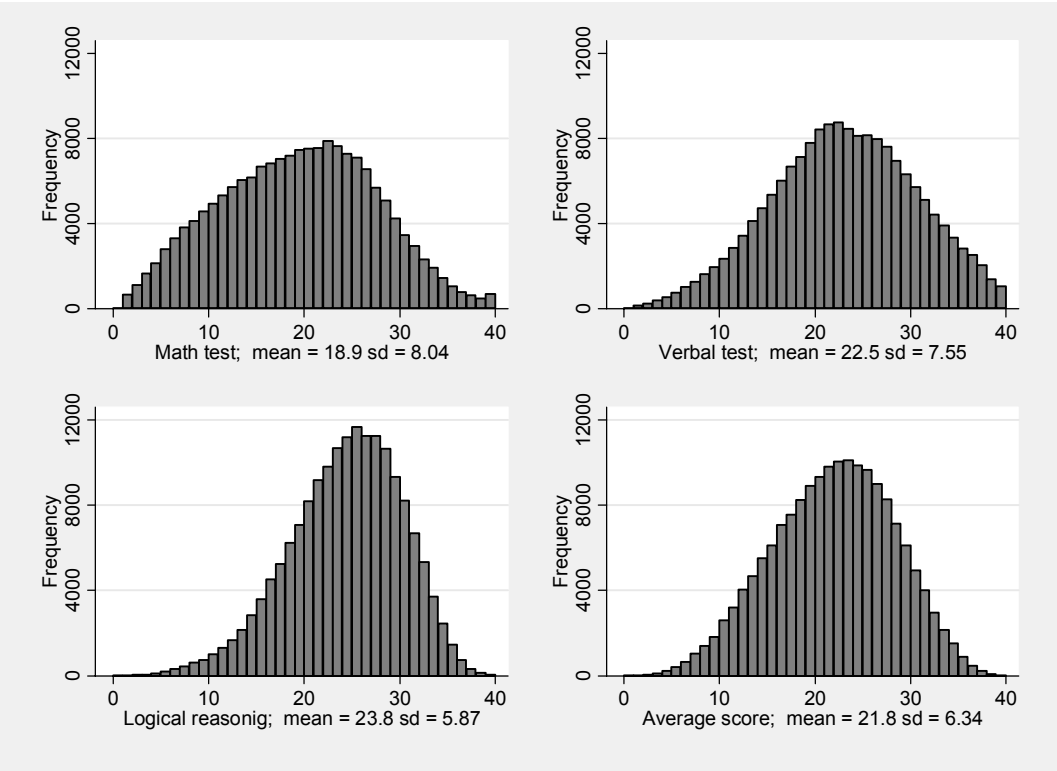
The test was originally created in 1955 and re-designed in 1981. Exactly the same test was used over the span of years that we analyze. From 1982 the test results are stored in the Army database that also includes personal identification numbers, making it possible to link the test results to information on test takers from other registers. Our data include all conscripts born between 1962 and 1966 who were found from the Army database, i.e. those who started their military service after January 1982. There is some selectivity in the data due to the fact that it

[2] The contents of the tests are described in detail in Tiihonen et al (2005).

is possible to enter to military service as a volunteer at age 17 and thus some men in the oldest cohorts served before the Army register was created. It is also possible to be exempted from the military service due to religious or ethical conviction, though in 1980s this was rare. The more common reasons for being exempt from military service are severe health conditions, most often related to mental health problems. However, even these criteria were substantially stricter in 1980s than what they are today. A comparison of the number of observations by birth cohort in our data and the corresponding cohort size in the 1984 population census reveals that our test score data contain information on 85.3 percent of the relevant male cohorts.

In Figure 3 we plot the distribution of the raw scores, i.e. the number of correct answers in each subtest. In the bottom right corner we plot the distribution of the average score. Two things might be noted from the Figure. First, there is plenty of variation in the test scores; the raw scores are distributed over the whole range from zero to forty. Second, the distribution of the test scores, particularly the distribution of the average score seems to be close to the normal distribution, a feature that we will exploit later.

Figure 3 Distribution of the test scores



Statistics Finland linked the test scores from the Army data to Census data on Finnish population. The Statistics Finland longitudinal census file contains data on the entire population living in Finland in 1970, -75, -80, -85 and -90. From 1990 onwards information is available for all years. Census data is mainly based on administrative registers. For example, information on education is based on the Register of Degrees and Examinations that collects data on all degrees granted directly from educational institutions. Data on place of residence in each census year is based on the Population Register. In general these register data are of very high quality. Only a few persons have any missing data, and the main reasons for not being included in the census data are residing abroad and death. In other words, our data does not suffer from attrition problems that often plague similar studies.

From the census data we gathered information on the date of birth and the place of residence in 1970, -75 and -80, which jointly determine whether the individual attended a tracked or a comprehensive school system. Statistics Finland does not release these data at a municipality-level, but per our request created an indicator classifying municipalities into six categories according to the year in which the comprehensive school reform was implemented in each municipality. Except for those who moved between census years between municipalities that implemented the reform at different years, we can accurately determine which school system was in place when the students were in the relevant age. The movers were dropped from the data used below.

The census data also include family codes that can used to identify brother pairs and to gather information on parents' education and earnings. To be more exact, these family codes are based on persons living in the same household, not necessarily biological family members. We use family codes from the 1975 census when the oldest men in the sample were 13 years old and most likely still living at home.

Table 1 reports the mean test scores by cohort and reform region. It shows that there are large differences across regions and a general increase in the test scores over time. These regional differences are correlated with the average education level and the average income in the region. An increase in the test scores over time, generally known as the Flynn-effect, has also been documented by Koivunen (2007) for a longer time period but naturally reflects also other differences between cohorts than those due to the school system.

The shaded area of the table indicates the students that went to the comprehensive school. Since these students are younger and concentrated in the regions with below average test scores, it is obvious that a cross-section comparison of regions or a time-series comparison of subsequent cohorts would not produce reliable estimates for the effect of the comprehensive school reform.

Table 1 Average test score by region and cohort

	Reform year						
Birth cohort	1972	1973	1974	1975	1976	1977	Total
1962	19.96	19.87	20.30	20.73	21.26	22.57	20.76
	[2,471]	[3,629]	[5,286]	[5,095]	[5,304]	[2,822]	[24,607]
1963	20.82	20.57	20.94	21.35	21.91	22.96	21.42
	[2,776]	[4,126]	[6,037]	[6,190]	[6,174]	[3,516]	[28,819]
1964	20.70	20.4	21.04	21.44	21.85	23.07	21.44
	[2,663]	[3,967]	[5,751]	[5,964]	[6,188]	[3,696]	[28,229]
1965	20.92	20.74	21.09	21.48	22.18	23.02	21.60
	[2,494]	[3,706]	[5,517]	[5,846]	[5,790]	[3,594]	[26,947]
1966	21.42	21.13	21.55	21.74	22.35	23.41	21.96
	[2,023]	[3,037]	[4,732]	[5,176]	[5,094]	[3,344]	[23,406]
Total	20.74	20.52	20.97	21.35	21.91	23.02	21.44
	[12,427]	[18,465]	[27,323]	[28,271]	[28,550]	[16,972]	[132,008]

Note: The shaded areas indicate cohorts that were affected by the post-reform educational system. Number of observations in each cell is in square brackets below the mean score in each cell.

4. Estimation methods

Our goal is to estimate the causal effect of the tracking regime on the army test scores. That is, we want to determine how an average student, or a student with certain characteristics, would have fared, had she or he been assigned to the reformed comprehensive system instead of the previous selective early tracking system. We use a fixed effects approach that controls for regional differences as well as general trends over time. The effect of the comprehensive school reform is identified because the timing of the reform differs across regions.

Most of our estimates are based on a following regression model:

$$y_{ijt} = \alpha + \Omega' D_{ij} + \Psi' D_{it} + \beta C_{ijt} + \varepsilon_{ijt} \qquad (1)$$

where y_{ijt} is the army test score of individual i who went to school in region j and belongs to cohort t. D_{ij} and D_{it} are region and cohort specific dummies, and C_{ijt} is an indicator for a pupil attending comprehensive school.

The parameter of interest in (1) is β. The identifying assumption is that the comprehensive school indicator, C_{ijt}, is uncorrelated with the error term conditional on the other regressors. This assumption, and the fact that D_i and D_{it} enter (1) additively, reflect the basic differences-in-differences assumptions. The parameter β is an unbiased estimate of the average causal effect of comprehensive schooling if the timing of the reform is uncorrelated with other region-specific changes in student outcomes.

In addition to the mean effects of the reform, we are also interested in the effect of the reform on the variance of the test scores. A natural way of examining this is the explicit modelling of the error variance of Equation (1). We could do this in two steps by taking the residuals from the Equation (1) and then explaining the squared residuals with the cohort and region dummies and the comprehensive school reform indicator. This procedure would be similar to standard heteroskedasticity tests. Since the OLS-residuals are unbiased estimates of the error terms, the two-step estimates are consistent.

A more straightforward way is to model simultaneously the effect of the reform on both the mean and the variance of the test scores. Assuming that the error term follows a normal distribution, the test scores will be distributed as

$$y_{ijt} \sim \frac{1}{\sqrt{2\pi\sigma_{jt}^2}} \exp\left[-\frac{1}{2} \frac{\left(y_i - (\alpha + \Omega D_{ij} + \Psi D_{it} + \beta C_{ijt})\right)^2}{\sigma_{jt}^2} \right]. \qquad (2)$$

The subscripts in σ_{jt}^2 indicate that the variance in the test scores may vary across regions and cohorts and may be affected by the reform. We parameterize the model assuming that log-variance is an additive function of the region, cohort and reform dummies. This also ensures that the variance estimates are positive and makes the coefficients easy to interpret.

$$\sigma^2{}_{jt} = \exp(\alpha + \Gamma D_j + \Phi D_t + \gamma C_{jt}) \qquad (3)$$

The log-likelihood function of the normal – heteroskedastic model is

$$\ln L = -\frac{N}{2}\ln(2\pi) - \frac{1}{2}\sum_{i=1}^{N}\left(\alpha + \Gamma D_{ij} + \Phi D_{it} + \gamma C_{ijt}\right) - \frac{1}{2}\sum_{i=1}^{N}\left[\frac{\left(y_{ijt} - \alpha - \Omega D_{ij} - \Psi D_{it} - \beta C_{ijt}\right)^2}{\exp\left(\alpha + \Gamma D_{ij} + \Phi D_{it} + \gamma C_{ijt}\right)}\right] \qquad (4)$$

where β measures the effect of the reform on the mean score and γ its effect on the variance.

5. Results

The baseline results are reported in Table 2. In Column (1) we simply regress the average test score on the comprehensive school dummy, and find that those who attended the comprehensive school scored on average 0.5 points less in the army test. However, the results in Column (2) reveal that this negative correlation reflects the fact that regions with lower test scores adopted the reform first. When full sets of birth cohort and region dummies are included in the regression, the effect of comprehensive school is close to zero. The effect becomes slightly larger but remains insignificant in Column (3) where we control for the age at the time of the test, and in Column (4) where we restrict the data to those who took the test at age 20.

It is not entirely clear whether controlling for the age at the test date is an appropriate approach. A fairly common reason for postponing military service is studying at a university. Also many high school graduates prefer serving in the military at age 19, before pursuing further studies at universities. This selectivity is reflected in our data. Those who take the test at age 20 tend to perform worse than those taking the test at age 19 or at age 21. On the other hand, limiting the analysis to the twenty-year-olds does correct for potential selectivity due to the fact that the test scores are available only from year 1982 onwards. In Column (5), we add family fixed effects to the equation, thus identifying the effect of the reform from the differences between brothers that attended different school systems. The effect remains small and insignificant. Interestingly, adding family fixed effects also reverses the positive trend in the test scores, indicating that the birth order effect on the test scores is larger than the difference across the birth cohorts.[3]

[3] The birth order effect was also found in a Norwegian study of the Army test scores (Kristenssen and Bjerkdal, 2007).

Table 2 Effects on test score average

	(1) No controls	(2) Region & cohort	(3) Age at test date	(4) Only 20-year-olds	(5) Family fixed effects
Reform	**-0.524*** **	**0.0009**	**0.062**	**0.056**	**0.111**
	(0.035)	**(0.059)**	**(0.057)**	**(0.066)**	**(0.146)**
Birth year 1963		0.549***	0.170***	0.153***	-0.196
		(0.054)	(0.053)	(0.059)	(0.133)
Birth year 1964		0.482***	0.187***	0.212***	-0.329**
		(0.058)	(0.057)	(0.064)	(0.139)
Birth year 1965		0.608***	0.386***	0.423***	-0.460***
		(0.065)	(0.063)	(0.071)	(0.157)
Birth year 1966		0.935***	0.495***	0.499***	-0.538***
		(0.071)	(0.070)	(0.080)	(0.174)
Reform region 1973		-0.203***	-0.211***	-0.273***	
		(0.070)	(0.068)	(0.078)	
Reform region 1974		0.222***	0.272***	0.247***	
		(0.066)	(0.064)	(0.074)	
Reform region 1975		0.527***	0.513***	0.561***	
		(0.069)	(0.067)	(0.078)	
Reform region 1976		1.139***	1.211***	1.286***	
		(0.074)	(0.072)	(0.084)	
Reform region 1977		2.111***	2.151***	2.552***	
		(0.086)	(0.084)	(0.098)	
Constant	21.96***	20.47***	19.94***	19.87***	21.11***
	(0.028)	(0.073)	(0.0713)	(0.081)	(0.0874)
Obs	142001	142001	142001	107930	107930
R-squared	0.002	0.015	0.074	0.020	0.002

Notes: The dependent variable is an unweighted average in three tests. In Column (3) 13 one year age dummies are included in the regression model, but not reported in the Table. Only 20 year old test takers are used in the family fixed effects regressions.

In Table 3 we examine separately the effect of the school reform on different tests. In Column (1) we regress each test score separately on the region and cohort dummies and a dummy variable indicating whether a person had attended a comprehensive school. In Column (2) we again add controls for age at the test date, in Column (3) limit the data to twenty-year-olds, and, in Column (4), add family fixed effects. For brevity we only report the coefficients of the comprehensive school dummy in each case. According to the results reported in Table 3, the comprehensive school reform had no significant effects on either math or logical reasoning tests. The effect on the verbal ability test is positive, but statistically significant only if we control for the age at the test date or limit the data to those who are 20 years old when taking the test. Even then the effect is small. The largest estimate, 0.172 points in the test, is only 0.02 standard deviations. The estimates that control for family fixed effects range from 0.05 in the verbal test to 0.16 in the logical reasoning test. However, family fixed effects estimates tend to be much less precise than the estimates that exploit also between-family variation, and

are therefore never significantly different from zero or significantly different from the point estimates reported in Columns (1)-(3).

The finding that the comprehensive school reform has its largest effects on the verbal test was perhaps to be expected. After all, verbal skills are something that are learned in schools, and so the changes in school system may well have effects on these verbal skills. If indeed the logical reasoning test truly measures innate reasoning abilities, pre-test schooling should have little or no effect on the test. Finally, the changes in the mathematics teaching resulting from the reform were perhaps not as significant. As noted above, the ability grouping was retained in mathematics and, as a result, math classes continued to be taught at three different ability levels after the reform.[4]

Table 3 Effects in different tests

	(1) Only region & cohort	(2) Age at test date	(3) Only 20-year-olds	(4) Family fixed effects
Math test	-0.063	0.014	0.022	0.102
	(0.074)	(0.072)	(0.083)	(0.191)
Verbal test	0.107	0.172**	0.143*	0.0658
	(0.070)	(0.068)	(0.078)	(0.178)
Logical reasoning	-0.011	0.032	0.035	0.161
	(0.054)	(0.054)	(0.062)	(0.152)

The entries in the table are coefficients of the dummy variable indicating that the person attended comprehensive school. Each regression model is estimated separately and includes cohort and region fixed effects. Only 20 year old test takers are used in the family fixed effects regressions

In Table 4 we report the maximum-likelihood estimates measuring the effects of the reform on both the mean and on the (log)variance of the test scores. We estimate these equations separately for each test. All equations include cohort and region effects on both the mean and the variance, but we report only the effects of the comprehensive school. According to Table

[4] As a robustness check we experimented by adding region-specific linear trends to the equations reported in Table 3. In these specifications the effect of the reform on the verbal test score gets somewhat larger; point estimate is 0.32. The effects on other two tests also grow slightly but remain statistically insignificant. We take these results as an indication that at least the qualitative results cannot be explained by pre-existing differences in trends in the test scores. The estimates are reported in the Appendix in the end of the paper. We also specified the model so that instead of treating the comprehensive school as a discrete indicator, we replaced it with a continuous variable that measures years of exposure to the comprehensive system. This varies between 5 and 9 years depending on whether the students entered the comprehensive system from the first grade or whether they were already in school when the reform was implemented in their area. These estimates do not suggest that the length of exposure matters, perhaps indicating that changes in tracking were more important than the changes in curriculum during first four years in school. Also these results can be found in the Appendix.

4 the maximum-likelihood method produces very similar estimates for the effect of the reform on the mean scores as the linear regression model used in Tables 2 and 3. The effects are significant only for the verbal test. The effects on the variance of the test scores are small. In the math test the effect is close to zero. In the verbal and logical reasoning test the reform reduced the variance between 1.7 and 2.6 percent. Only the effect on the logical reasoning test is statistically significant. Since the test score data that we use are the raw scores from an identical test implemented for different cohorts, issues such as standardization of the test scores or variation in the test across the years cannot explain the results.

Another natural way of examining the effects of the reform on the distribution of the test scores would be to use quantile regression models. However, even though we measure test scores with a 40 point scale, we cannot detect effects on quantiles, unless the effect is at least one full point. Since even the largest observed effects are below 0.2 points, a quantile regression did not produce meaningful results.

Table 4 ML-estimates of the effect of the reform on mean and variance of the test scores

	(1) Math	(2) Verbal	(3) Logical reasoning	(4) Average score
Effect on mean	0.012	0.168**	0.039	0.062
	(0.072)	(0.068)	(0.053)	(0.057)
Effect on log variance	-0.002	-0.016	-0.026**	-0.017
	(0.013)	(0.013)	(0.013)	(0.013)

In Tables 5A and 5B we examine the effects of the comprehensive school reform by family background. In Table 5A we estimate regression models similar to those reported in Column (2) of Table 3 but add an indicator of parents' education and its interaction with the reform dummy. We classify parents as having higher education if at least one of the parents has completed at least 12 years of education. In the pre-reform schooling system this generally refers to a situation where the parent attended the more academic track. The parents' income is measured by summing the annual taxable income of both parents, deflating the income to the 1980 price level and taking an average over the census years 1970, -75 and -80.

According to the results reported in Table 5A, parental schooling has a clear effect on the test scores. Men with highly educated parents have 2.0 points higher score in the verbal test, 2.2

points higher score in the math test, and 1.5 points higher score in the logical reasoning test. The effect of the reform – now referring to the effect on those with less educated parents – is positive and statistically significant in the verbal test and positive but insignificant in the other two tests. More importantly, the interaction between parents' education and the comprehensive schooling is always negative and in all but the logical reasoning test statistically significant. A closer look at the average test score reveals that the reform increased the score for those with less educated parents by about 0.2, but the interaction coefficient is approximately equally large, producing a zero effect for men with highly educated parents.

Table 5B repeats the analysis using parents' income. The results are qualitatively similar to those with parents' education. Men with richer parents tend to score better in all tests, and the interaction between the parents' income and the reform dummy is negative for all tests.

Table 5A Effect of the reform by parents' education

	(1) Math	(2) Verbal	(3) Logical Reasoning	(4) Average score
High ed. parents	2.013*** (0.0734)	1.826*** (0.0692)	1.355*** (0.0545)	1.733*** (0.058)
Reform	0.180** (0.0894)	0.306*** (0.0843)	0.125* (0.0664)	0.196*** (0.070)
Reform × high ed. parents	-0.281*** (0.0905)	-0.213** (0.0854)	-0.157** (0.0672)	-0.224*** (0.071)
Constant	11.84*** (0.563)	16.58*** (0.531)	19.67*** (0.417)	16.11*** (0.443)
Observations	127 206	127404	127432	127 167
R-squared	0.043	0.072	0.059	0.092

Table 5B Effect of the reform by parents' income

	(1) Math	(2) Verbal	(3) Logical Reasoning	(4) Average score
Parents' income	2.465*** (0.0698)	2.105*** (0.0659)	1.614*** (0.0519)	2.048*** (0.055)
Reform	0.0397 (0.0753)	0.202*** (0.0712)	0.0470 (0.0560)	0.085 (0.059)
Reform × parents' income	-0.275*** (0.0844)	-0.212*** (0.0797)	-0.168*** (0.0628)	-0.212*** (0.066)
Constant	13.19*** (0.559)	17.80*** (0.528)	20.58*** (0.415)	17.27*** (0.440)
Observations	127 120	127 318	127 346	127 081
R-squared	0.094	0.079	0.066	0.102

6. Conclusions

Persistent differences in average test scores across countries and over time have received plenty of attention in recent years. One often suggested explanation for these differences is the educational system. In particular, the tracking of pupils into different groups by ability and aspirations has been considered a potentially important factor. However, both the economic theory and the available empirical evidence remain inconclusive when it comes to the effects of tracking regimes on test scores.

In this paper we have estimated the effect of the comprehensive school reform on the Finnish Army Basic Skills Test scores. Unlike previous literature that had to rely on cross-country comparisons or comparisons of regions within countries, we can estimate the effect of the comprehensive school reform on test scores using a differences-in–differences approach and single-country data. As such, our study provides a more serious attempt at identifying the causal effect of school systems on test outcomes.

We find that the reform had a small positive effect on the average verbal test scores and no significant positive or negative effect on the average arithmetic or logical reasoning test results. The effect on the verbal test scores appears to arise entirely from families where the parents had only basic education. In this group the reform had a positive effect also on the test scores in the arithmetic or logical reasoning tests. Finally, we find that the effect of the reform on the variance of the test scores is very small.

Overall, our results indicate that the effects of school tracking on cognitive skills tend to be small. The important effects of tracking on other later-in-life outcomes such as earnings, completed schooling or college enrolment that have been reported in previous research are therefore likely to be related to other mechanisms than the direct effects on skill formation.

References

Aakvik, A. – Salvanes, K. G. – K. Vaage, (2003): "Measuring heterogeneity in the returns to schooling in Norway using educational reforms", Centre for Economic Policy Research, Discussion paper No. 4088.

Black, S.E. – P. Deveraux – K. Salvanes. (2005): "Why the apple doesn't fall far: Understanding the intergenerational transmission of education", *The American Economic Review* 95, pp. 437-449.

Brunello, Giorgio – Daniele Checchi (2007): "Does school tracking affect equality of opportunity? New international evidence", *Economic Policy*, Oct 2007, pp. 781-861.

Brunello, Giorgio – Massimo Giannini – Kenn Ariga (2004): "The optimal timing of school tracking", IZA Discussion Paper No. 995.

Duflo, Esther – Pascaline Dupas – Michael Kremer (2008): "Peer effects and the impact of tracking: Evidence from a randomized evaluation in Kenya", NBER Working Paper, 14475.

Galindo-Rueda, Fernando – Anna Vignoles (2005): "The heterogeneous effect of selection in secondary schools: Understanding the changing role of ability," CEE Discussion Paper 52.

Hanushek, Eric – Ludger Wößmann (2006): "Does educational tracking affect performance and inequality? Differences-in-differences evidence across countries," *Economic Journal* 116, C63-C76.

Holmlund, Helena (2008): "Intergenerational mobility and assortative mating: Effects of an educational reform", CEE Discussion Paper 91.

Hoxby, M.C. (2001): "Peer effects in the classroom: Learning from gender and race variation", NBER working paper 7867.

Kerckhoff, Alan C. – Ken Fogelman – David Crook – David Reeder (1996): "*Going Comprehensive in England and Wales. A Study of Uneven Change,* London: The Woburn Press.

Koivunen, Salla (2007): Suomalaismiesten kognitiivisen kykyprofiilin muutokset 1988-2001. Flynnin efektiä suomalaisessa aineistossa?[Changes in cognitive skill profile among Finnish men. Flynn effect in Finnish data?], Master's thesis, University of Jyväskylä.

Kristensen Petter – Tor Bjerkedal (2007): "Explaining the Relation Between Birth Order and Intelligence", *Science*, 316 (5832), 1717.

Lazear, Edward, P. (2001): "Educational Production", *The Quarterly Journal of Economics,* Vol. 116, No. 3 (Aug., 2001), pp. 777-803.

Manning, Alan – Jörn-Steffen Pischke (2006): "Comprehensive versus selective schooling in England in Wales: What do we know?" NBER Working Paper No. 12176.

Meghir, C. – M. Palme, (2005): ”Educational reform, ability, and parental background”, *American Economic Review*, 95 (1), 414-424.

Pekkarinen, T. – R. Uusitalo – S. Pekkala (2006): Education policy and intergenerational income mobility: Evidence from the Finnish comprehensive school reform, IZA Discussion Paper 2204.

Statistics Finland (1986): “Structure of population and vital statistics”, Official Statistics of Finland VI A:150, Central Statistical Office of Finland, Helsinki 1986.

Tiihonen, J. – J. Haukka – M. Henriksson – M. Cannon – T. Kieseppä – I. Laaksonen – J. Sinivuo – J. Lönnqvist, (2005): ”Premorbid intellectual functioning in bipolar disorder and schizophrenia: Results from a cohort study of male conscripts”, *American Journal of Psychiatry*, 162, 1904-1910.

Waldinger, Fabian (2006): “Does tracking affect the importance of family background on students' test scores?” mimeo, London School of Economics.

Appendix: Additional results

Table A1 Effect by years in comprehensive school

VARIABLES	(1) Math	(2) Verbal	(3) Logical Reasoning	(4) Average score
5 years in comprehensive	0.021	0.205**	0.001	0.063
	(0.105)	(0.099)	(0.0788)	(0.0831)
6 years in comprehensive	-0.048	0.123	0.120	0.057
	(0.142)	(0.134)	(0.107)	(0.112)
7 years in comprehensive	-0.175	0.153	0.052	-0.001
	(0.186)	(0.175)	(0.139)	(0.147)
8 years in comprehensive	-0.106	0.186	0.127	0.057
	(0.237)	(0.223)	(0.177)	(0.187)
9 years in comprehensive	-0.040	0.317	0.001	0.079
	(0.307)	(0.289)	(0.230)	(0.242)
Constant	16.69***	20.53***	22.33***	19.87***
	(0.173)	(0.163)	(0.129)	(0.136)
Observations	107974	108176	108208	107930
R-squared	0.016	0.014	0.018	0.020

In all columns the model includes a full set of dummy variables for region and cohort

Table A2 Allowing for linear regional trends in test scores

VARIABLES	(1) Math	(2) Verbal	(3) Logical Reasoning	(4) Average score
Reform	0.113	0.317***	0.061	0.153**
	(0.088)	(0.083)	(0.065)	(0.069)
Constant	12.65***	16.98***	20.17***	16.66***
	(0.522)	(0.492)	(0.387)	(0.411)
Observations	142051	142286	142322	142001
R-squared	0.070	0.058	0.047	0.074

In all columns the model includes a full set of dummy variables for region and their interactions with a linear time trend as well as a full set of cohort dummies.

IN VATT WORKING PAPERS SERIES PUBLISHED PUBLICATIONS

1. Tomi Kyyrä – Pierpaolo Parrotta – Michael Rosholm: The effect of receiving supplementary UI benefits on unemployment duration. Helsinki 2009.

VALTION TALOUDELLINEN TUTKIMUSKESKUS
STATENS EKONOMISKA FORSKNINGSCENTRAL
GOVERNMENT INSTITUTE FOR ECONOMIC RESEARCH

Valtion taloudellinen tutkimuskeskus
Government Institute for Economic Research
P.O.Box 1279
FI-00101 Helsinki
Finland

ISBN 978-951-561-841-2
ISSN 1798-0283