School tracking and development of cognitive skills

by

Sari Pekkala (Charles River Associates International)

Tuomas Pekkarinen (IFAU, Uppsala University)[1]

and Roope Uusitalo (Helsinki School of Economics)

Abstract

This paper estimates the effect of a major education reform on test scores that measure cognitive skills. The Finnish comprehensive school reform of 1972-1977 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 by imposing a uniform curriculum on all the pupils. We estimate the effect of the reform on Finnish Army Basic Skills tests that measure arithmetic, verbal, and visuopspatial skills. Our data contain the whole population of conscripts who take the tests when they are 19-21 old. The identification strategy relies on differences-in-differences strategy and exploits the fact that the reform was implemented gradually across the country during a six-year period. We find that the reform had a small positive effect on the verbal test scores. This effect was strongest in families where parents had only basic education.

[1] Corresponding author. Adress: IFAU, Box 513, 751 20 Uppsala, Sweden. Email: tuomas.pekkarinen@ifau.uu.se.

1. Introduction

The publication of the results from the OECD Programme for International Student Assessment (PISA) 2003 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 from other countries with roughly equal school resources. In addition, there seems to be no clear efficiency-equity trade-off. For example, Finnish students were top performers in the most recent PISA study. Still also the variance in test scores turned out to be among the smallest in Finland.

If different resources do not explain the cross-country differences in achievement, the explanation must be related to something about the way school systems are organized. Several authors have argued that a partial explanation has to do with tracking or ability grouping of students. Theoretically, ability grouping might be beneficial for learning since it allows schools to target teaching to more homogeneous groups. However, if there are substantial peer-effects the students that are placed to lower-ability tracks may suffer from ability grouping. Overall benefits depend on whether peer-effects are linear or non-linear (Hoxby, 2001) An additional complication arises from difficulties in measuring ability at an early age. Errors in measurement lead to students being placed on a wrong track. Track placement may also be affected by family background (Brunello, Giannini, and Ariga, 2004) From a theoretical point of view it is therefore not at all clear how tracking should affect the mean and variance of educational performance.

The issue is not empirically settled either. For example, Hanushek and Wössmann (2006) use data from international student assessments and find that early tracking increases inequality in student achievement. At the same time early tracking seems to have generally negative effects on mean performance, though the evidence on mean effects is less consistent. The Hanushek Wössmann study is based on cross-country differences in school systems. A similar cross-country approach is used by Brunello and Checchi (2006) and Waldinger (2006) who both study whether family background is more important for student achievement in countries where tracking occurs early. Neither of these studies confirms that tracking would increase inequality in achievement though Brunello and Checchi find that early tracking enforces the effect of family background on early wages.

Conflicting results from previous studies partly reflect the difficulties in analysing the effect of tracking based on cross-country data. While these studies try to control for variation due to other

factors by including early test scores (Hanushek and Wossmann, 2006; and Waldinger, 2006) or by using time variation in tracking age (Brunello and Checchi, 2006) it is far from clear that all relevant cross-country differences would be accounted for.

An ideal test of the tracking effects occurs when a country switches from tracked to comprehensive system. Effects of such reforms have been evaluated recently by Manning and Pischke (2006) for the UK and Maurin and McNally (2007) for Northern Ireland. Even in these cases it is often difficult to separate regional differences from the effects of school systems, particularly if only cross-section or single cohort data is available. For example, Manning and Pischke note that areas that switched to comprehensive system first were on average poorer than areas that retained the tracked system. They also reach a conclusion that even instrumental variables approach that uses arguably exogenous variation in the timing of the reform is unlikely to solve the selectivity problem.

In this paper we use data from Finnish comprehensive school reform implemented between 1972 and 1977 to evaluate the effects of tracking on formation of cognitive skills. The Finnish reform replaced the old two-track school system with nine-year comprehensive school and imposed a uniform curriculum on the entire cohort up to age 16. The key feature of the reform was that it was implemented gradually starting from the northernmost parts of the country and spreading to south during a period of six years. This gradual implementation allows controlling for regional variation and time trends using a difference-in-differences approach. Furthermore our data also include information on families which makes possible to control for the family effects and to estimate the effects of the reform using differences across brothers that were placed to different school systems.

2. Comprehensive school reform [2]

2.1 Background

Finland introduced a wide-ranging comprehensive school reform in the 1970's. Similar reforms had already taken place in Sweden and Norway (Meghir and Palme, 2004; Aalvik, Salvanes and Vaage, 2003). The comprehensive school reform abolished the old two-track school system with a uniform

[2] For a more detailed description of the Finnish comprehensive school system and the reform see Pekkarinen, Uusitalo and Pekkala (2006).

comprehensive school lasting for nine years. 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 (civic schools) that kept almost the whole 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 finished their schooling. The pre-reform system is described schematically in the left-hand panel of Figure 1.

[FIGURE 1 SCHOOL SYSTEMS]

2.2 Content of the comprehensive school reform

The school system was reformed during the 1970s. 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. Students were also to select courses in subjects such as handwork and additional languages which was argued to improve the individuality of the education. 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 post-reform 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 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 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 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 when the region was supposed to start implementing the reform according to the regional implementation plan. After this, each incoming cohort 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 so that all the pupils in the grades 1-9 were in the comprehensive school.

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 – COMPREHENSIVE SCHOOL REFORM MAP]

2.4 The comprehensive school reform as a quasi-experiment

The Finnish comprehensive school reform is in many ways an ideal experiment for evaluating the effects of early versus late tracking on the cognitive outcomes. The regional implementation plan 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 remain some caveats to the approach. First of all, as is clear from Figure 2, there were exceptions to the geographical implementation plan. Some municipalities implemented the reform earlier that the rest of the municipalities in the region. 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 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.

3. Data

A fundamental problem in assessing the effects of a school reform on student performance is that students in separate school systems rarely participate in comparable tests. One possibility is to use nation-wide or international comparisons on 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 with a reasonable sample size implemented to students in both pre- and post reform school systems.

For this paper we managed to get hold of basic skills test results from the Finnish Army. Since military service is mandatory in Finland the almost the entire male cohort is tested. The Army test is given to all new conscripts between the third and sixth week of their service. Average age at the time when these men are tested is 20, so clearly 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 immediate effects on test results. The Finnish Army test is also a strong predictor of earnings 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 has been designed to measure general ability and logical thinking. The test consists of three subtests: verbal, arithmetic, and visuospatial. Each test includes 40 questions sorted in increasing difficulty. The aggregate scores from the test are used in selecting conscripts to officer training. Descriptive statistics on the raw test scores are presented in table 1 and their distributions in figures 3a to 3c.

[TABLE 1: DESCRITIVE STATISTICS ON THE TEST SCORES]

[FIGURES 3a-3c: TEST SCORE HISTOGRAMS]

Especially the verbal and arithmetic reasoning tests test that kind of skills that should be affected by the quantity and the quality of the pre-test schooling. In the verbal reasoning subtest, the subject has to choose synonyms or antonyms of give words, to select words that belonging to the same category as a given word pair, to exclude words from a group of works, and to choose similar relationships between word pairs. The arithmetic reasoning test asks the subject to complete series, solve verbally expressed problems, compute simple arithmetic operations, and to choose similar relationships between pairs of numbers. The visuospatial reasoning test is based on Raven's progressive matrices and should be viewed as a standard "culture free" intelligence test.[3]

The test was created in 1955 but re-designed in 1981. Exactly the same test was used over a span of years that we analyze. From 1982 the test results are included in the Army database that also includes personal identification numbers that can be linked to other register data. Our data include

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

on all conscripts born between 1962 and 1966 that were found from the Army database. There is some selectivity in the data due to the fact that it is possible to enter to military service as a volunteer at age 18 and some men in the oldest cohorts served before the Army register was created. It is also possible to be exempted from the service due to religious or ethical conviction, but in 1980s this was extremely rare. More common is exemption due to severe health conditions such as diabetes of mental health problems. Still the comparison of number of observation by birth cohort and the corresponding cohort size in the 1984 census data reveals that data contain information on 85.3 percent of the relevant male cohorts.

The Army data was linked at Statistics Finland to census data on Finnish population. The Statistics Finland longitudinal census file contains data on the entire population living in Finland at each census year 1970, -75, -80, -85 and -90. From 1990 onwards data is collected every year. Information is mostly based on administrative registers. For example income data comes from tax registers, information on the level of education is based on Register of Degrees and Examinations and labour market status is based on pension insurance registers and Ministry of Labour register on unemployed job seekers. Data on place of residence in each census year is based on Population Register. In general these register data is of very high quality. Only few persons have missing data and only reasons of not being included in the census data arte death and emigration.

From the census data we gathered information on the dates of birth and region of residence in 1970, -75 and -80 which determine whether the individual attended tracked or comprehensive school system. Statistics Finland does not release these data on an individual municipality level, but for our request created an indicator classifying municipalities according to the year when 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 could accurately determine which school system was in place when they were in relevant age. Those who moved were dropped from the analysis data.

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. Finally, the data contain information on own education and earnings of the conscripts recorded every census year from 1970 to 2000.

For the purposes of this paper, we restricted the sample to those individuals that took their army tests when they were between 19 and 21 olds. This is the time when most conscripts go to army and selectivity is likely to be a problem among those who take the test earlier and later. This restriction leaves us with information on 132 330 individuals of whom 132 053 took the arithmetic test, 132 282 took the verbal test, and 132 313 took the visuospatial test.

Table 2 describes how the sample is divided into different cohorts and across the reform regions. There are no large differences in the cohort size in these age groups. The most intense reform years were 1974, 1975, and 1976. The table also shows how the treatment status depends on birth year and timing of the reform in the municipality of residence. In the 1962 cohort, for example, only those who went to school in the regions that adopted the comprehensive school in 1972 or 1973 were affected by the reform. The shaded area in the table indicates that there are a number of potential difference-in-differences estimates that can be calculated to evaluate the effect of the reform.

[TABLE 2: TIMING OF THE REFORM BY COHORT]

4 Estimation methods

Our goal is to estimate the causal effect of the tracking regime on 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 comprehensive system instead of the old early tracking system. Apart from this, we are also interested in the distributional effects of the late-tracking comprehensive system. We would, for example, like to know whether the variation of test scores decreased as a result of transition to comprehensive system and whether the effect of the reform on test scores was stronger at the lower or upper tail of the test score distribution.

The problem with analysing the causal effects of the comprehensive school on test scores is that pupils select themselves to tracking regimes based on some unobservable individual or family characteristics. As was explained at the outset, the previous literature has tried to tackle this problem by comparing individuals in different countries or within countries that are in the process of switching from one system to another.

The strategy chose here is similar to the latter approach used also by Manning and Pischke (2006) as well as Maurin and McNally (2007). However, unlike these authors who had to rely on data on a single cohort and assume that there are no systematic differences in ability or family background characteristics across regions, we use data on multiple cohorts. Hence, we can control for both region and cohort specific fixed effects and estimate the effect of the comprehensive school by differences in differences.

Our regression model is as follows:

$$y_{ijt} = \alpha + \Omega' D_j + \Psi' D_t + \beta C_{jt} + \Gamma' F_i + \varepsilon_{ijt} \quad (1)$$

where y_{ijt} is the army test score of individual i who went to school in municipality j and belongs to cohort t. D_j and D_t are region and cohort specific dummies, F_i is a set of proxies for family background, and C_{jt} is an indicator for a pupil attending comprehensive school.

The parameter of interest in (1) is β. The identifying assumption is that comprehensive school indicator, C_{jt}, is uncorrelated with the error term conditional on the other regressors, ie. $E(\varepsilon_{ijt} \mid C_{jt}, D_j, D_t, F_i) = E(\varepsilon_{ijt} \mid D_j, D_t, F_i)$. This assumption and the fact that D_j and D_t enter (1) additively reflect the basic differences-in-differences assumption: that is, we assume that the change in y_{ijt} in a region that switches to comprehensive system between cohorts t and $t-1$, would have been the same as in other regions in the absence of the change in school system.

Below, we will include various proxies for family background F_i, such as parental education. However, since it is unlikely that these proxies will capture all the effects of family background and since our data allow us to identify families, we will also estimate the effect of tracking regime on test scores, by exploiting within siblings variation in school systems. This can be clarified by writing separate regression models for younger and older brothers:

$$y_{i,j,young} = \alpha + \Omega' D_j + \Psi' D_{young} + \beta C_{j,young} + \Gamma' F_i + \varepsilon_{i,j,young} \quad (2)$$

and

$$y_{i,j,old} = \alpha + \Omega' D_j + \Psi' D_{old} + \beta C_{j,old} + \Gamma' F_i + \varepsilon_{i,j,old} \quad (3)$$

In the families where younger brother attended the comprehensive system and the older brother went into the old school system, the within siblings difference in test scores will identify the effect of comprehensive system on test scores conditional on fixed family effects:

$$y_{young,j,t} - y_{old,j,t} = \Psi'(D_{young} - D_{old}) + \beta(C_{j,young} - C_{j,old}) + \varepsilon_{i,j,young} - \varepsilon_{i,j,old} \quad (4)$$

In regression (4), all the unobservable family characteristics that are fixed in time will be controlled for. Below, we will estimate the effect of the comprehensive school on test score with both regressions (1) and (4). In order to explore the distributional effects of the comprehensive school, we will also interact the effect of comprehensive school with parental education and the average test scores in the family.

5 Results

In tables 3 to 5, we first report the average test scores by reform regions and birth cohorts. The first column of the upper panel in tables 3 to 5 displays the estimate by birth cohort. There is a clear indication of an upward trend in these test scores. Test scores increase by approximately one point across these five cohorts. This has also been documented by Koivunen (2007) for a longer time period and may also reflect other differences between cohorts than the schooling system. In the second and third columns, we report the average test scores separately in the regions where the reform had not taken place by the time the when the cohort turned eleven and in regions where the system was already reformed. The rightmost column reports the within-cohort difference between these regions. The differences are all significant and indicate that regions that adopted the reform earlier had clearly lower test scores.

The bottom panels of tables 3 to 5 repeats these calculations now examining differences over birth cohorts within regions. The first columns show that there are substantial differences across regions. In the second and third column, we report the average test scores separately for the pre- and post-reform cohorts. In all the regions all the test scores are higher for the younger cohorts that went through the comprehensive system.

[TABLE 3: ARITHMETIC TEST SCORE RESULTS BY REGION AND COHORT]

[TABLE 4: VERBAL TEST SCORE RESULTS BY REGION AND COHORT]

[TABLE 5: VISUOSPATIAL TEST SCORE RESULTS BY REGION AND COHORT]

Tables 6 to 8 report the results from the estimation of equation (1). The first column of each table reports the simple differences in means between pre- and post-reform regions. The test scores are always significantly lower in the comprehensive system. However, the results in the second column 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 comprehensive school did not have a significant effect on arithmetic or visuopspatial test scores. However, the reform had a small positive effect, 0.143 points, on the verbal test score.

In the fourth column of tables 6 to 8, we add a dummy that takes value one if one of the parents of the tested individual has a degree from at least a secondary school. This has the effect of increasing the effect of comprehensive school reform on verbal test scores slightly. When we estimate the equation (1) separately for families where parents have only basic education, in column (4), and for families where one of the parents has at least a secondary degree, in column (5), we find that the effect on verbal test scores comes solely from families with basic education. The results in tables 6 to 8 indicate that the effects of the reform are sensitive to introduction of controls for family background variables. Although our data set would allow for a rich set of these variables, it seems more reasonable to control for fixed family effects by estimating the effect of the reform on within sibling differences in test scores.

In tables 9 to 11, we report the results from the estimation of equation (4). The differences are calculated as the differences between younger and older brothers. This means that the mean differences in test scores are always negative. The effect of the reform is identified by families where the younger brother attended a comprehensive school and the older brother attended pre-reform school. Thus, a positive effect of the reform should show up as a positive coefficient on the difference in the reform dummies.

The first columns report the simple reform effect on the difference. This is never significant and introducing controls for the age difference does not change this pattern. However, when we interact the reform dummy with a dummy for parents with secondary or higher degree, the reform seems to have a marginally positive main effect on all the test scores. Once again this implies that the effect of the reform was strongest in the families with only basic education. The effect is still very small however. The results indicate that in the families where parents had only basic education, the reform increased test scores by one third of a point.

In order to examine the effect of the reform at different ability levels, we interacted the reform dummy with the average test score of the brothers. The results from this exercise are reported in the fourth column of tables 9 to 11. The main effect now reports the effect on the mean test score and is never significant. However, the interaction with the average test score is negative and at least close to significant in all the test scores. This relationship is the strongest with the verbal test scores. If one calculates the marginal effect of the reform on verbal test scores by different percentiles of the brothers' average test score, one finds that the effect is 0.70 points at the 10^{th} percentile while it is clearly negative at the 90^{th} percentile. The same pattern is repeated with the rest of the test scores as well although the relationship is not as strong.

6 Conclusions

Persistent differences in average test scores across countries and time have received a lot 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 seen as an important factor. However, both the economic theory and available empirical evidence is 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. The comprehensive school reform changed the Finnish educational system from an early tracking system into a comprehensive, late tracking system. Unlike previous literature that has had to rely on cross-country comparisons of comparisons of regions within countries, we can estimate the effect of the comprehensive school reform on test scores by differences in differences. Since the reform was adopted at different times in different parts of the country, we are able to control for both region and cohort specific fixed effects, when estimating the effect of the reform.

We find that the reform had a modest positive effect on the average verbal test scores and no significant positive or negative effect on arithmetic or visuospatial test results. The effect on the verbal test scores comes entirely from families where parents had only basic education. When we control for family fixed effects by focusing on differences within siblings, we find that the reform had a small positive effect on all test scores in the families where parents had only basic education. Furthermore, it seems that the effect of the reform was strongest at the lower tail of the ability

distribution. The effect on the verbal test score, for example, is twice as large as the mean effect at the 10th percentile of the brothers' average test score distribution.

References

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

Brunello, Giorgio and Daniele Checchi (2006): "Does School Tracking Affect Equality of Opportunity? New International Evidence", IZA Discussion Paper No. 2348

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

Hanushek, Eric and Ludger Wößmann (2006): "Does Educational Tracking Affect Performance and Inequality? Differences-in-Differences Evidence across Countries," *Economic Journal* 116, C63-C76.

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

Koivunen, Salla (2007): *Suomalaismiesten kognitiivisen kykyprofiilin muutokset 1988-2001. Flynnin efektiä suomalaisessa aineistossa?*, Master's thesis, University of Jyväskylä.

Manning, Alan and Jörn-Steffen Pischke (2006): "Comprehensive versus Selective Schooling in England in Wales: What Do We Know?" NBER Working Paper No. 12176

Maurin, Eric and Sandra McNally (2007): "Educational Effects of Widening Access to the Academic Track: A Natural Experiment" IZA Discussion Paper No. 2596

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

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, and 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

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

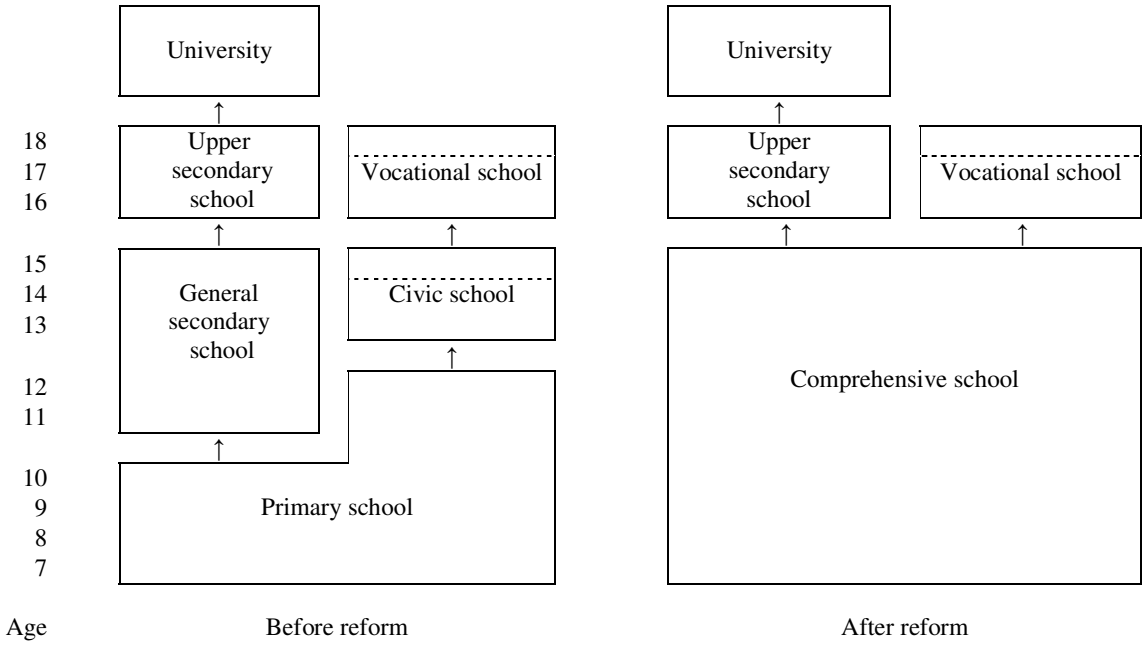

Figure 2 The implementation of the comprehensive school reform across regions 1972-1977

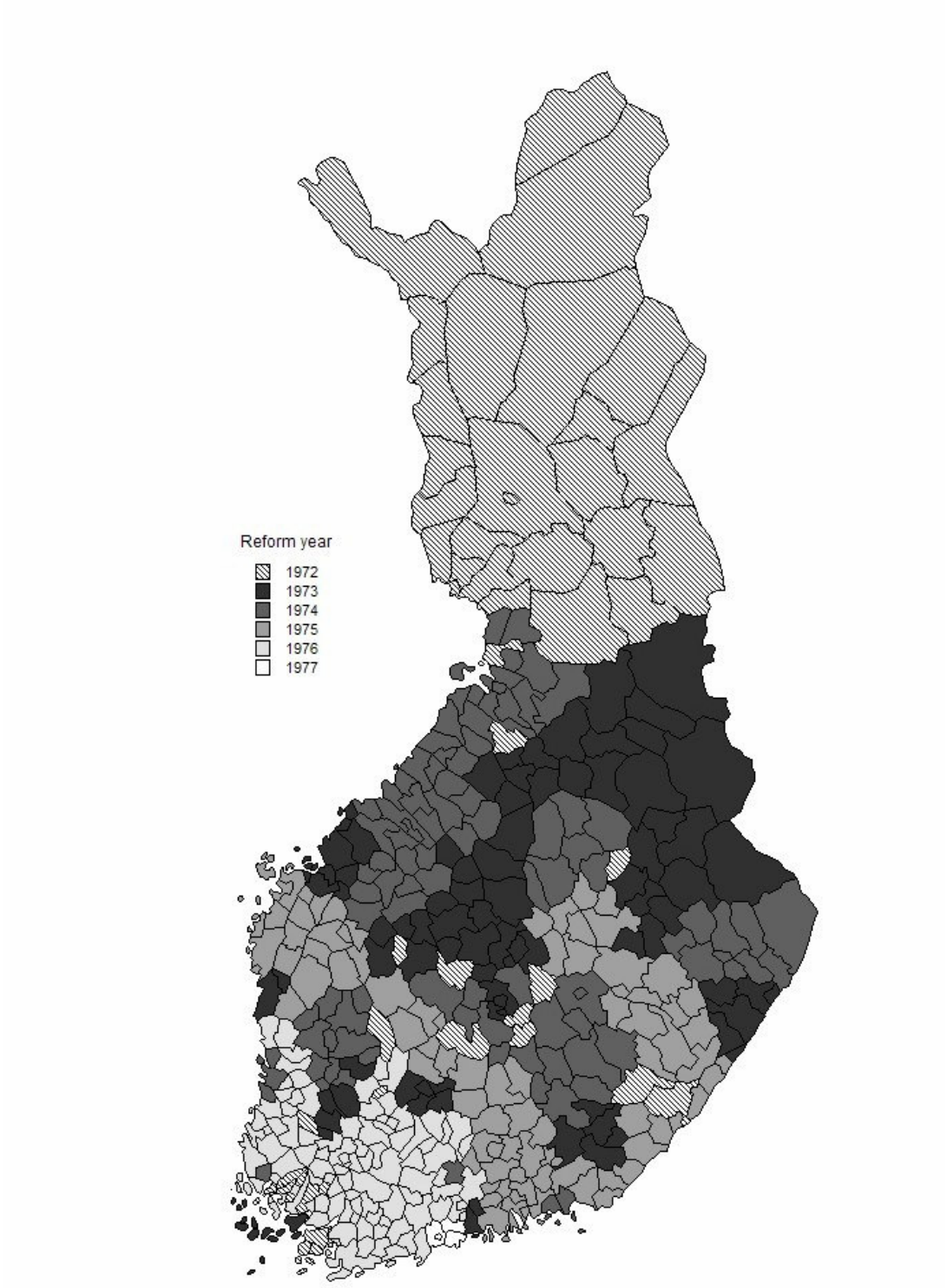

Figure 3a Histogram of the raw verbal test scores in the whole population

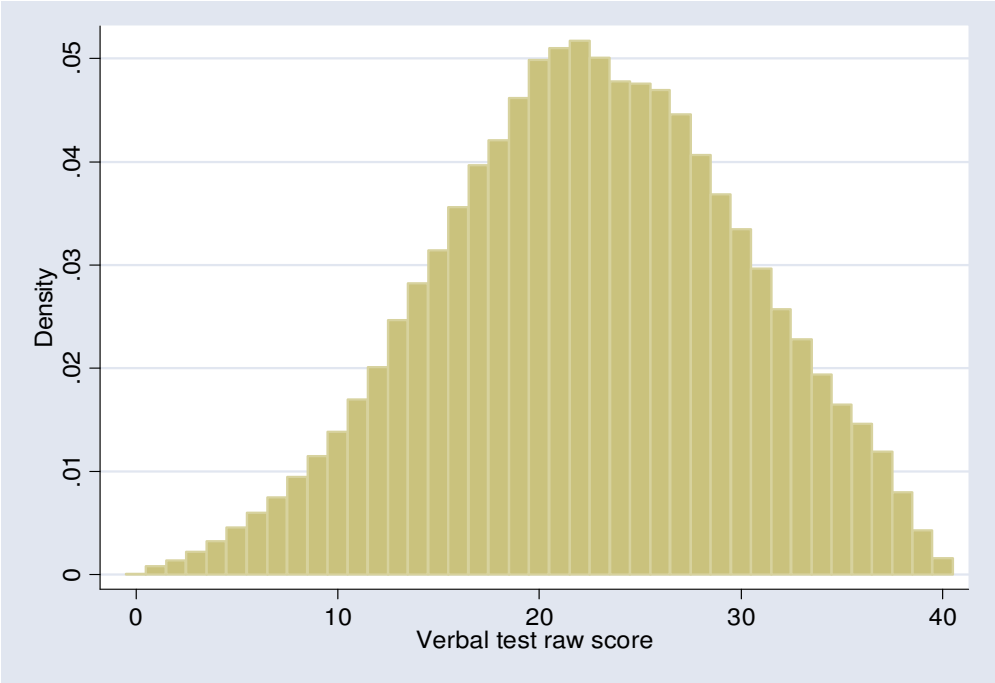

Figure 3b Histogram of the raw arithmetic test scores in the whole population

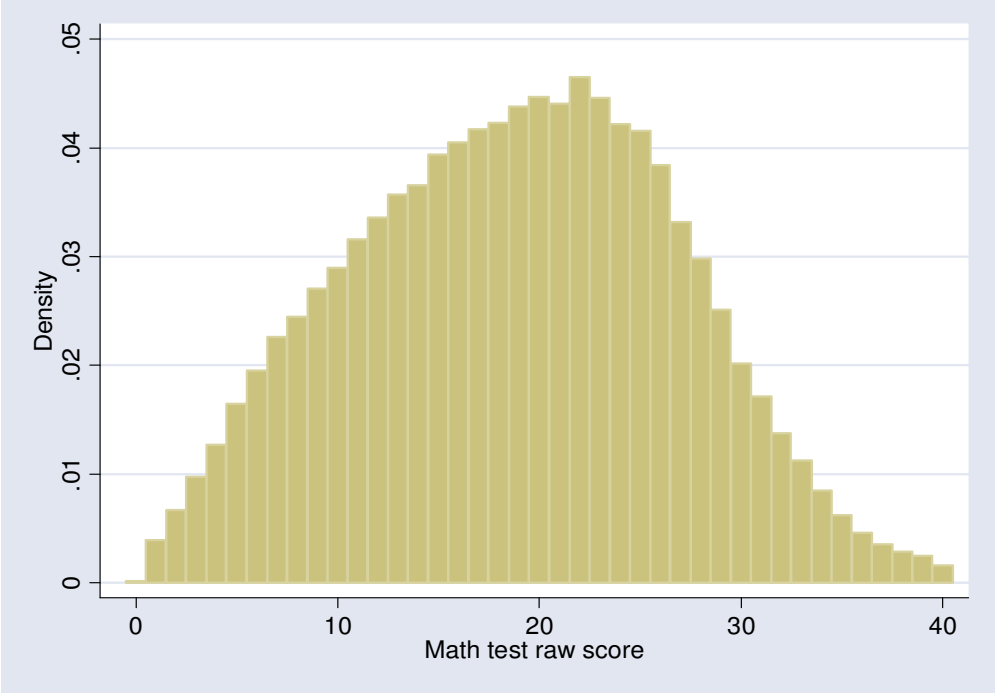

Figure 3c Histogram of the raw visuospatial test scores in the whole population

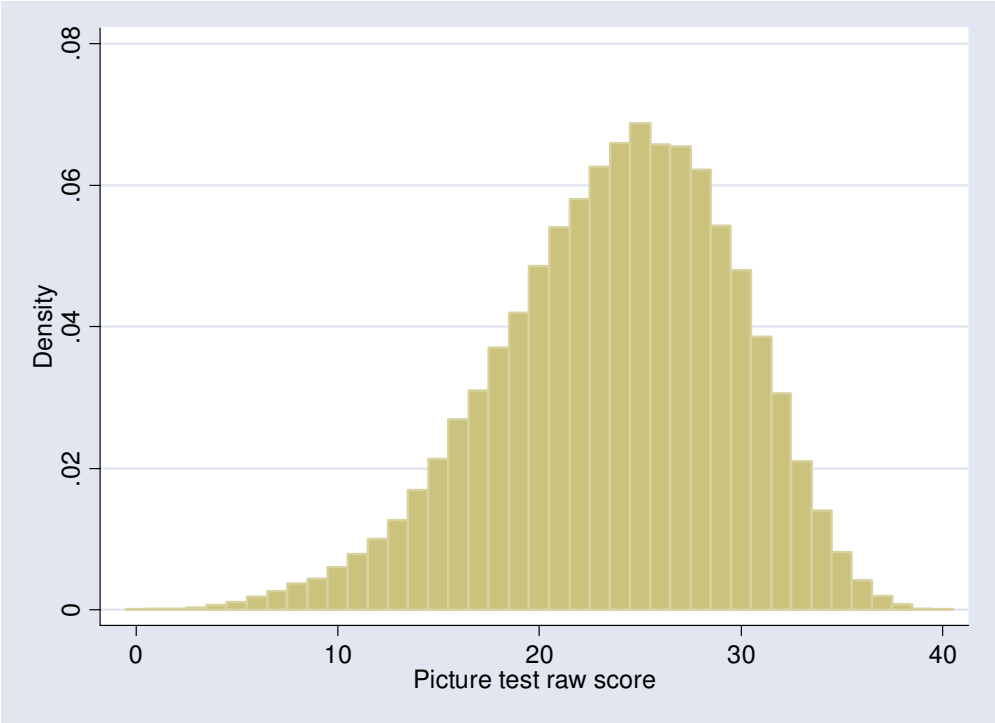

Table 1 Summary statistics on the test scores

Variable	Obs	Mean	Std. Dev.	Min	Max
Verbal	149 870	22.50	7.51	0	40
Arithmetic	149 617	18.86	8.03	0	40
Visuospatial	149 908	23.76	5.87	0	40

Table 2 The timing of the reform by cohorts and regions

	Reform year						
Birth cohort	1972	1973	1974	1975	1976	1977	Total
1962	4th grade	5th grade	6th grade	7th grade	8th grade	9th grade	
	N=2,476	N=3,637	N=5,306	N=5,117	N=5,323	N=2,829	N=24,688
1963	3rd grade	4th grade	5th grade	6th grade	7th grade	8th grade	
	N=2,779	N=4,146	N=6,052	N=6,208	N=6,189	N=3,522	N=28,896
1964	2nd grade	3rd grade	4th grade	5th grade	6th grade	7th grade	
	N=2,668	N=3,980	N=5,766	N=5,975	N=6,200	N=3,707	N=28,296
1965	1st grade	2nd grade	3rd grade	4th grade	5th grade	6th grade	
	N=2,498	N=3,716	N=5,527	N=5,858	N=5,800	N=3,606	N=27,005
1966	-	1st grade	2nd grade	3rd grade	4th grade	5th grade	
	N=2,028	N=3,042	N=4,738	N=5,189	N=5,103	N=3,345	N=23,445
Total	N=12,449	N=18,521	N=27,389	N=28,347	N=28,615	N=17,009	N=132,330

Note: The shaded areas indicate cells that adopted the post-reform educational system. N refers to the sample size in each cell in the data that are used in the analysis.

Table 3 Arithmetic test scores across birth cohorts and reform regions

a) Birth cohorts

Birth cohort	Average	Pre-reform	Post-reform	Difference
1962	17.708	18.025	16.744	-1.281
	(0.050)	(0.057)	(0.099)	(0.114)
1963	18.489	19.087	17.754	-1.332
	(0.047)	(0.063)	(0.069)	(0.093)
1964	18.485	19.501	17.938	-1.563
	(0.048)	(0.081)	(0.059)	(0.099)
1965	18.623	20.345	18.358	-1.987
	(0.048)	(0.132)	(0.051)	(0.141)
1966	19.082		19.082	
	(0.052)		(0.052)	

b) Reform regions

Region	Average	Pre-reform	Post-reform	Difference
1972	17.585		17.585	
	(0.070)		(0.070)	
1973	17.469		17.469	
	(0.058)		(0.058)	
1974	17.933	17.206	18.107	0.901
	(0.047)	(0.107)	(0.053)	(0.120)
1975	18.416	18.089	18.634	0.545
	(0.047)	(0.074)	(0.061)	(0.096)
1976	19.028	18.837	19.336	0.499
	(0.047)	(0.059)	(0.076)	(0.096)
1977	20.264	20.157	20.700	0.543
	(0.061)	(0.068)	(0.139)	(0.153)

Note: Numbers in the cells mean arithmetic test scores. Standard errors are reported in parentheses.

Table 4 Verbal test scores across birth cohorts and reform regions

a) Birth cohorts

Birth cohort	Average	Pre-reform	Post-reform	Difference
1962	21.486	21.764	20.642	-1.122
	(0.047)	(0.054)	(0.097)	(0.109)
1963	22.182	22.642	21.618	-1.024
	(0.044)	(0.059)	(0.066)	(0.088)
1964	22.224	23.010	21.801	-1.209
	(0.045)	(0.074)	(0.056)	(0.093)
1965	22.391	23.636	22.199	-1.437
	(0.045)	(0.122)	(0.048)	(0.132)
1966	22.708		22.708	
	(0.048)		(0.048)	

b) Reform regions

Region	Average	Pre-reform	Post-reform	Difference
1972	21.566		21.566	
	(0.066)		(0.066)	
1973	21.224		21.224	
	(0.056)		(0.056)	
1974	21.827	21.077	22.007	0.930
	(0.045)	(0.102)	(0.050)	(0.113)
1975	22.072	21.728	22.301	0.573
	(0.044)	(0.070)	(0.057)	(0.091)
1976	22.676	22.428	23.079	0.651
	(0.043)	(0.055)	(0.068)	(0.088)
1977	23.717	23.621	24.108	0.487
	(0.056)	(0.063)	(0.124)	(0.141)

Note: Numbers in the cells mean verbal test scores. Standard errors are reported in parentheses.

Table 5 Visuospatial test scores across birth cohorts and reform regions

a) Birth cohorts

Birth cohort	Average	Pre-reform	Post-reform	Difference
1962	23.019	23.253	22.308	-0.945
	(0.038)	(0.043)	(0.076)	(0.087)
1963	23.522	23.988	22.952	-1.036
	(0.034)	(0.046)	(0.052)	(0.069)
1964	23.572	24.353	23.152	-1.202
	(0.035)	(0.057)	(0.044)	(0.073)
1965	23.733	24.984	23.540	-1.444
	(0.035)	(0.094)	(0.038)	(0.103)
1966	24.059		24.059	
	(0.038)		(0.038)	

b) Reform regions

Region	Average	Pre-reform	Post-reform	Difference
1972	23.037		23.037	
	(0.053)		(0.053)	
1973	22.797		22.797	
	(0.044)		(0.044)	
1974	23.106	22.550	23.240	0.690
	(0.036)	(0.081)	(0.040)	(0.090)
1975	23.511	23.311	23.645	0.334
	(0.035)	(0.055)	(0.044)	(0.071)
1976	23.967	23.754	24.315	0.561
	(0.033)	(0.043)	(0.053)	(0.069)
1977	25.034	24.939	25.424	0.485
	(0.043)	(0.048)	(0.094)	(0.108)

Note: Numbers in the cells mean visuospatial test scores. Standard errors are reported in parentheses.

Table 6 Regression results – arithmetic test scores

	(1)	(2)	(3)	(4) Low educated parents	(5) High educated parents
Reform	-0.598	-0.044	-0.023	0.081	-0.148
	(0.045)	(0.076)	(0.079)	(0.115)	(0.102)
Birth year 1963		0.768	0.702	0.584	0.860
		(0.070)	(0.073)	(0.103)	(0.095)
Birth year 1964		0.747	0.641	0.508	0.871
		(0.076)	(0.079)	(0.113)	(0.101)
Birth year 1965		0.889	0.755	0.612	1.025
		(0.085)	(0.088)	(0.127)	(0.113)
Birth year 1966		1.322	1.156	0.958	1.482
		(0.093)	(0.097)	(0.137)	(0.126)
Reform region 1973		-0.118	-0.027	-0.045	-0.110
		(0.091)	(0.094)	(0.128)	(0.128)
Reform region 1974		0.329	0.336	0.237	0.387
		(0.086)	(0.089)	(0.124)	(0.119)
Reform region 1975		0.788	0.819	0.824	0.698
		(0.090)	(0.093)	(0.131)	(0.124)
Reform region 1976		1.397	1.371	1.319	1.317
		(0.097)	(0.101)	(0.143)	(0.132)
Reform region 1977		2.599	1.772	1.833	2.377
		(0.113)	(0.119)	(0.182)	(0.147)
Highest years of education in the household			0.473		
			(0.007)		
Constant	18.856	16.903	11.748	16.185	17.621
	(0.036)	(0.096)	(0.127)	(0.138)	(0.131)
Observations	132053	132053	117919	56885	75168
R-squared	0.00	0.01	0.05	0.01	0.01
Standard errors in parentheses					

Note: The dependent variable is the raw test score in arithmetic reasoning test. Reform refers to the comprehensive school reform dummy. Cohort dummies refer to 5 birth cohort dummies that are included in the regression. Region dummies refer to 6 reform region dummies that are included in the regression. Highest years of education in the household is the maximum of the parents' years of education. Low educated parents refer s to families where the highest education is 9 years long. Higher educated parents have more than 9 years of education.

Table 7 Regression results – verbal test scores

	(1)	(2)	(3)		
				Low educated parents	High educated parents
Reform	-0.402	0.143	0.173	0.204	0.095
	(0.043)	(0.072)	(0.075)	(0.108)	(0.095)
Birth year 1963		0.649	0.566	0.506	0.709
		(0.066)	(0.069)	(0.097)	(0.089)
Birth year 1964		0.638	0.546	0.462	0.713
		(0.071)	(0.074)	(0.106)	(0.095)
Birth year 1965		0.769	0.664	0.636	0.794
		(0.080)	(0.083)	(0.119)	(0.106)
Birth year 1966		1.039	0.858	0.834	1.076
		(0.088)	(0.091)	(0.130)	(0.118)
Reform region 1973		-0.344	-0.288	-0.282	-0.329
		(0.086)	(0.088)	(0.121)	(0.120)
Reform region 1974		0.281	0.282	0.218	0.311
		(0.081)	(0.084)	(0.117)	(0.112)
Reform region 1975		0.542	0.554	0.604	0.431
		(0.085)	(0.088)	(0.124)	(0.116)
Reform region 1976		1.184	1.154	1.139	1.086
		(0.091)	(0.095)	(0.135)	(0.123)
Reform region 1977		2.230	1.479	1.581	2.000
		(0.106)	(0.112)	(0.172)	(0.138)
Highest years of education in the household			0.429		
			(0.007)		
Constant	22.453	20.818	16.139	20.098	21.532
	(0.034)	(0.090)	(0.120)	(0.130)	(0.123)
Observations	132282	132282	118112	56992	75290
R-squared	0.00	0.01	0.05	0.01	0.01
Standard errors in parentheses					

Note: The dependent variable is the raw test score in verbal reasoning test. Reform refers to the comprehensive school reform dummy. Cohort dummies refer to 5 birth cohort dummies that are included in the regression. Region dummies refer to 6 reform region dummies that are included in the regression. Highest years of education in the household is the maximum of the parents' years of education. Low educated parents refer s to families where the highest education is 9 years long. Higher educated parents have more than 9 years of education.

Table 8 Regression results – visuospatial test scores

	(1)	(2)	(3)	(4) Low educated parents	(5) High educated parents
Reform	-0.433	-0.001	0.010	0.081	-0.056
	(0.033)	(0.056)	(0.059)	(0.086)	(0.074)
Birth year 1963		0.487	0.431	0.354	0.551
		(0.052)	(0.054)	(0.077)	(0.069)
Birth year 1964		0.516	0.429	0.348	0.595
		(0.056)	(0.058)	(0.084)	(0.074)
Birth year 1965		0.673	0.576	0.514	0.731
		(0.062)	(0.065)	(0.095)	(0.082)
Birth year 1966		0.975	0.848	0.821	1.002
		(0.069)	(0.072)	(0.103)	(0.092)
Reform region 1973		-0.242	-0.216	-0.244	-0.181
		(0.067)	(0.070)	(0.096)	(0.093)
Reform region 1974		0.062	0.042	0.046	0.060
		(0.064)	(0.066)	(0.093)	(0.087)
Reform region 1975		0.455	0.439	0.548	0.337
		(0.067)	(0.069)	(0.098)	(0.090)
Reform region 1976		0.916	0.871	0.913	0.824
		(0.072)	(0.075)	(0.107)	(0.096)
Reform region 1977		1.962	1.412	1.566	1.752
		(0.083)	(0.088)	(0.136)	(0.107)
Highest years of education in the household			0.293		
			(0.005)		
Constant	23.853	22.524	19.369	21.971	23.065
	(0.027)	(0.070)	(0.094)	(0.104)	(0.095)
Observations	132313	132313	118136	57010	75303
R-squared	0.00	0.02	0.04	0.01	0.02
Standard errors in parentheses					

Note: The dependent variable is the raw test score in visuospatial reasoning test. Reform refers to the comprehensive school reform dummy. Cohort dummies refer to 5 birth cohort dummies that are included in the regression. Region dummies refer to 6 reform region dummies that are included in the regression. Highest years of education in the household is the maximum of the parents' years of education. Low educated parents refer s to families where the highest education is 9 years long. Higher educated parents have more than 9 years of education.

Table 9 Within sibling regressions – arithmetic test results

	(1)	(2)	(3)	(4)
Sibling reform difference	0.117	0.077	0.375	0.090
	(0.153)	(0.166)	(0.218)	(0.166)
Sibling age difference		0.050	0.050	0.051
		(0.080)	(0.083)	(0.080)
Parents' education x reform difference			-0.521	
			(0.258)	
Mean arithmetic score x reform difference				-0.034
				(0.019)
Constant	-0.353	-0.430	-0.432	-0.432
	(0.090)	(0.154)	(0.160)	(0.154)
Observations	13490	13490	12441	13490
R-squared	0.00	0.00	0.00	0.00
Standard errors in parentheses				

Note: The dependent variable is the difference in the arithmetic test score between younger and older brothers. Sibling reform difference is the difference in reform dummies. Age difference is the difference in birth years. Parents' education is a dummy for parents with a secondary or higher degree. Mean arithmetic score is the brothers' mean score.

Table 10 Within sibling regressions – verbal test results

	(1)	(2)	(3)	(4)
Sibling reform difference	0.163	0.155	0.350	0.172
	(0.143)	(0.155)	(0.203)	(0.155)
Sibling age difference		0.009	-0.006	0.013
		(0.075)	(0.077)	(0.074)
Parents' education x reform difference			-0.243	
			(0.241)	
Mean verbal score x reform difference				-0.063
				(0.019)
Constant	-0.588	-0.602	-0.607	-0.608
	(0.084)	(0.144)	(0.150)	(0.143)
Observations	13535	13535	12482	13535
R-squared	0.00	0.00	0.00	0.00
Standard errors in parentheses				

Note: The dependent variable is the difference in the verbal test score between younger and older brothers. Sibling reform difference is the difference in reform dummies. Age difference is the difference in birth years. Parents' education is a dummy for parents with a secondary or higher degree. Mean verbal score is the brothers' mean score.

Table 11 Within sibling regressions – visuospatial test results

	(1)	(2)	(3)	(4)
Sibling reform difference	0.152	0.173	0.265	0.184
	(0.121)	(0.131)	(0.172)	(0.131)
Sibling age difference		-0.026	-0.006	-0.025
		(0.063)	(0.065)	(0.063)
Parents' education x reform difference			-0.165	
			(0.203)	
Mean verbal score x reform difference				-0.035
				(0.020)
Constant	-0.170	-0.130	-0.204	-0.131
	(0.071)	(0.122)	(0.127)	(0.122)
Observations	13544	13544	12489	13544
R-squared	0.00	0.00	0.00	0.00
Standard errors in parentheses				

Note: The dependent variable is the difference in the verbal test score between younger and older brothers. Sibling reform difference is the difference in reform dummies. Age difference is the difference in birth years. Parents' education is a dummy for parents with a secondary or higher degree. Mean verbal score is the brothers' mean score.