Supporting Information

Brinch and Galloway 10.1073/pnas.1106077109

SI Text

Further Details on Reform. In our analysis, we need to know whether each birth cohort in any given municipality experienced the old (prereform) or new (postreform) educational system. Written documentation on the timing of the reform, including a list by Ness (1) as well the series of official statistics (2–6), is limited and/or incomplete, and therefore not sufficient to pinpoint the exact timing of the reform in all the relevant municipalities. However, it is feasible to determine the timing of the reform if we have data of sufficient quality on place of residence at the relevant age (14 y) and educational attainment. Previous studies (7–10) have used a similar strategy but relied primarily on census data (from 1960) on place of residence; such information is of varying quality as an indication for place of residence at the relevant age for the different birth cohorts affected by the reform. In addition, municipality structure underwent a number of major changes in the early 1960s in Norway; thus, a place of residence recorded in the 1960 census may have ceased to exist as an independent municipality just a few years later. As a result, these previous studies would not have been able to establish place of residence as accurately at the relevant age for all cohorts or to pinpoint the timing of the reform with as much precision as in this current study. This does not, by any means, discredit the results from those studies, because the IV/2SLS approach accounts for such possible errors. However, data of poorer quality do lead to more statistical uncertainty (in the form of larger SEs), which, in turn, makes it more difficult to reject any hypotheses of no effect. Data of better quality allow us to estimate relevant relationships with greater precision.

For each male in the Norwegian population from birth cohorts for the period 1950–1958, we obtain information on the place of residence (municipality) at the time when he could have entered the eighth grade in the new educational system (in practice, January 1 of the year he turned 14 y of age). This information from the population register is matched with the educational register, which provides us with the individual's highest level of complete education at the age of 30 y. Data from earlier ages (i.e., before 1980 for our cohorts) are of questionable quality. We are able to observe that before the reform, a substantial fraction of a given birth cohort did not have education at the middle-school level and that their schooling is characterized by codes referring to the old school system [i.e., *folkeskole* (primary school), *framhaldsskole*, *realskole*]. Following the introduction of the reform, any relevant postreform birth cohort would be expected to have a substantial portion of individuals with educational codes referring to the new school system (i.e., completed *ungdomsskole*). By plotting the share of persons with old and new education codes at the primary- and middle-school levels over cohorts for each municipality, we are generally able to pinpoint the timing of the reform in each municipality. It is worthwhile to note that we are using exact (6-digit) educational codes for type of school diploma rather than simply years of completed education when we pinpoint the timing of the reform. In our work, the timing of the reform is usually very clear, because the old schooling codes simply cease to exist [i.e., the fraction with those types of educational codes drops to (nearly) 0] for residents in a given municipality.

Institutional factors hindered the identification of reform timing in many of the municipalities. During the time period we study, a large consolidation of municipality structure took place in Norway and resulted in a number of municipalities merging and others splitting up, mostly in 1964 and 1965. If two municipalities A and B merged to form municipality C at a time when municipality A had already implemented the reform but municipality B had not, we are unable to assign a reform year to municipality C and we are forced to exclude all observations of individuals from municipalities A and B. In other words, many merged municipalities where the reform was implemented in different years in different parts are left out of our reform year dataset. Alternatively, if municipality D splits into municipalities E and F and municipality G splits into municipalities E and G, municipalities C, D, E, F, and G can be treated as a single unit as long as the reform timing in each separate presplit municipality is compatible. About 13% of the pupils in the relevant years are left out of the sample for such reasons. We were also forced to exclude the capital city, Oslo, from the analysis because of inconsistencies in the data from the relevant period. Official statistics (3–6) list a large number of the new type of middle school (*ungdomsskole*) in Oslo many years before official reform introduction, as reported in other sources (1). (Results from analyses performed with observations from Oslo differed little from the other results presented here.)

For a number of municipalities, it is still difficult to assign a specific year of reform introduction even with access to detailed individual education data and despite our best efforts. In very small municipalities (of which there were many in Norway at the time), random variation in schooling choices involving just a handful of pupils in any given year would be enough to obscure any systematic change in schooling patterns. Furthermore, given the isolated location of many municipalities in Norway, a number of local idiosyncrasies in the educational system did exist at that time. For example, before the reform, youths from many small rural municipalities would have attended postcompulsory school in a larger neighboring municipality. Thus, when an old *realskole* offering education to a number of neighboring municipalities is turned into a new *ungdomsskole* as a result of the reform in the municipality where the old *realskole* was located, education in the new school may be offered to inhabitants in neighboring municipalities, but mandatory 9-y education only applies to the municipality where the school is located. The full introduction of the reform (with mandatory 9-y schooling) in the other involved municipalities may have occurred later. Such issues made it difficult for us to pinpoint the timing of the reform for ~18% of the individuals relevant for this study. Altogether, we are able to assign reform years to the municipalities where ~60% of the birth cohorts during the period 1950–1958 lived at the age of 14 y (Table S3). Table S1 provides descriptive statistics on the full sample of men in the 1950–1958 cohorts as well as the men who resided in municipalities for which we could identify the timing of the reform (i.e., the sample of analysis). It documents that the sample used in the analysis (i.e., for whom reform timing could be identified) did not differ markedly from the general population at the time.

There is some indication of partial treatment of the last prereform cohort in the data (e.g., Figs. S1 and S2). Indeed, there are several reasons why we should expect to find some sign of partial treatment of the last prereform cohort in our analysis. The series of official statistics (2–6) discusses and documents how correspondence between birth cohort and school cohort is imperfect, because parents could apply to have their child start first grade later than the norm (which was the fall of the year the child turned 7 y old) and school progression was delayed for some students because of sickness and/or grade retention. Roughly 5% of a birth cohort started school a year late, and roughly another

5% experienced delays in school progression (i.e., grade retention). Furthermore, we cannot dismiss the possibility that some students, who had completed the old (prereform) seventh grade a year earlier, returned to school as new middle schools opened nearby. Indeed, there is no reason to suppose that older students were explicitly barred from receiving further education in the new middle schools once they were established in a given municipality. Also, because many of these communities did actually have to build and/or create new middle schools to implement the reform fully, it seems likely that there would have been interest in filling up the new middle schools.

Altogether, we need to address partial treatment in the last prereform cohort, and we have three options for doing so, all of which were covered in this analysis and reported either in the main text or here. The main results exclude the last prereform cohort from the analysis. Table 2 also reports results with the last prereform cohort considered untreated. Note that inclusion of the last prereform cohort as untreated results in attenuation of the estimates (i.e., makes it more difficult to uncover a reform effect, because the last prereform cohort was partially treated). Finally, further results in which the last prereform cohort is considered fully treated (i.e., the reform is predated by 1 y) are also reported and briefly discussed below.

Further Details on Data. The data on IQ are taken from the draft assessment of the Norwegian military. Only in extreme circumstances, such as severe handicap, are men exempt from the preliminary draft assessment, of which the cognitive ability test is a part. Sorting of men for military duty, including further decisions about exemption from military service for health reasons, occurs after the cognitive ability test is taken. The test consists of three timed subtests: arithmetic, word similarities, and figures. The results from the subtests are combined into a general ability (GA) score standardized to a stanine (9-point) scale; the correlation between GA and Wechsler Adult Intelligence Scale IQ is 0.73 (11). Further details on these tests can also found elsewhere (12–14). In line with the common practice used in the studies cited above, we convert the stanine scores on GA into the more common IQ scale, with a mean of 100 and an SD of 15. Major changes in the subtests occurred either before or after the period relevant for our study (13). We were only able to obtain the GA data for persons born 1950 onward as part of this study.

As in most previous studies using the Norwegian cognitive ability data (e.g., 13, 15), we convert the stanine scores from the GA test of the Norwegian military to IQ equivalent scores by setting the stanine score of 5 equal to –100 and then using increments of 7.5 for each stanine score deviation from 5. This is the conversion based on standardization from the 1954 draft cohort (i.e., persons born about 1935). Because of the rise in IQ in Norway, the mean IQ for the cohorts we study is therefore considerably higher than 100 and the SD is lower (because of ceiling effects). Unfortunately, cognitive ability results broken down into the three subtests (arithmetic, word similarities, and figures) are not available for the full sample of cohorts we study.

The data in the NUDB starts with self-reported data on highest level of completed education from the 1970 census. Since 1974, relevant information on participation in education and completion of degrees has been reported directly by the educational institutions and not by the individuals themselves. As a result of "missing" data for the years 1971–1973, data on the highest level of education are only reliable after corrections and updating could be made following the census in 1980. We therefore use information on the highest level of education at the age of 30 y; this would be in 1980 for the earliest cohort we study (born in 1950).

Complementary Analyses. Fig. S2 relates the share of persons with less than 9 y of education to reform timing for municipalities that introduced the reform in the years 1952–1957 and helps us to document a number of relevant insights for understanding the introduction of the compulsory schooling reform. (The reform years 1952–1957 are chosen so as to have observations for cohorts both before and following the first full reform cohorts in Fig. S2.) The first insight provided by Fig. S2 is that the reform, as expected, dramatically lowered the share of persons with less than 9 y of education. Furthermore, we can see from Fig. S2 that ∼15–20% of the pupils were affected by the reform for each reform year studied. In other words, the reform did not simply have an impact on a very small and highly select group of pupils but altered education attainment for a substantial minority, almost one in five, of the youth population at the time. Finally, Fig. S2 documents the start of the decline of persons with less than 9 y of education in the year before full reform introduction (i.e., going from −2 to −1 on the horizontal axis, as discussed above).

The DID estimates presented in the main text implicitly assume that the effect of the reform does not vary over time after (or before) the reform. We can test this assumption by estimating a model that allows for variation in the reform effect over time to reform, as presented in Table S4. To allow sufficient flexibility, we include several periods well before or after any periods for which we might expect to find an effect. There is some evidence to suggest partial treatment of the last prereform cohort. Therefore, we include indicator variables for each of 4 y before the reform plus an aggregate category for ≥5 y before the reform. We do not suspect any differential effects over the years after the reform; however, to provide sufficient flexibility, we include indicator variables for each of 3 y postreform (i.e., 0, 1, 2 y after the reform) plus an aggregate category for ≥3 y following the reform. The reference category for time before/after indicators is the first postreform cohort (i.e., 0 y after the reform). The results in Table S4 are in line with all the previously reported results. The coefficients on the prereform categories for time to reform are roughly the same as the reform effects we estimate for the DID in the main text. The postreform categories do not differ among themselves, and none of the reported postreform coefficients are individually significantly different from the reference category. Thus, there is little evidence to suggest that the effect of the reform varied over the time leading up to or following the reform, with the one exception of the last prereform cohort, as previously reported.

Table S5 presents results in which we include the last prereform cohort as treated by simply predating the reform by 1 y in each municipality. Note that because compulsory eighth- and ninth-grade schooling did not apply to the last prereform cohort, this exercise has the potential of introducing a bias in who is treated in the last prereform cohort. For example, it is reasonable to assume that some of the most motivated youths in the last prereform cohort chose to pursue further education voluntarily in the new middle schools. If this is the case, we would expect a lower effect of the reform on education and higher (biased) estimate of the effect of education on IQ. The DID results in Table S5 show hardly any difference for the effect of the reform on IQ compared with Table 1 in the main text, whereas the IV/2SLS strategy suggests a lesser effect of the reform on education and a larger effect of education on IQ compared with Table 1 in the main text.

Flynn Effect in Norway. Norway was one of the countries discussed in Flynn's seminal paper (12) documenting a widespread rise in IQ over many years in a wide range of countries; the Norwegian data from the military draft are also judged to be quite comprehensive and of particularly high quality. However, Flynn (12) only had access to certain sporadic years for Norway in his paper; his documentation suggests that Norway experienced quite a dramatic rise of 10 IQ points between the 1954 and 1968 draft cohorts (which corresponds to birth cohorts during the period 1935–1949), followed by what appears to be leveling off in the subsequent 10–12 y. A more recent study by Sundet et al. (13)

provides further details and more comprehensive documentation starting from the 1954 draft cohort (born about 1935) through the 2002 draft cohort (born roughly about 1983). That study presents sporadic results for individual draft cohorts from 1954 until the end of the 1960s and then a long consecutive time series starting from the late 1960s through 2002.

Taking the two end points in the period for the analysis (13) suggests a total gain in IQ, relative to the 1954 mean of 100, of 10.8 IQ points, or an average of 0.23 points per year from 1954 to 2002. However, the results of the study by Sundet et al. (13) clearly indicate a much larger increase in average IQ before the period relevant for our cohorts [i.e., before (draft year) 1969]. The rise in the period relevant for our cohorts (1969–1977) is steady until a noticeable decline occurs toward the end of the 1970s. The start of that decline is also apparent in our data, for the birth cohorts in 1967 and 1968 (Fig. S1). The analysis in Sundet et al. (13) also indicates that that decline was reversed around the start of the 1980s, when scores started to rise steadily again before another decline, and leveling off occurred starting in the mid-1990s. The long time series (13) therefore suggest that increases in IQ have moved somewhat up and down around a rough but clear upward trend over the full period they study; the largest increases did clearly occur before (draft year) 1969. The study by Sundet et al. (13) also suggests that ceiling effects may have depressed the increase in scores as average IQ levels rise.

If we take the average IQ from the two end points in our data, we observe an increase of 1.56 IQ points from the 1950 birth cohort to the 1958 birth cohort, or an average annual rate of increase of ~0.2 points (Table S1). An estimated Flynn effect for the cohorts we study of ~0.202 points per year, based on a regression with a linear time trend and municipal indicator variables, is reported in the main text. Although Sundet et al. (13) do not report average IQs in a table or explicitly calculate average annual rates specifically for the period relevant for our study, the average annual increases we observe in our data are roughly similar to what appears in figure 1 in Sundet et al. (13) for the draft cohorts corresponding to our birth cohorts.

One further point to note in comparing our descriptive statistics with those of Sundet et al. (13) is that those researchers add 2.1 points to the scores for draft cohorts 1969–2001 to account for changes in the test. All such changes took place before or after the period we study (13). We therefore do not explicitly have to account for them in our study, but that will have an effect on comparisons of reported averages or levels across different studies.

1. Ness E, ed (1971) *Yearbook of the School* (Johan Grundt Tanum Forlag, Oslo) (in Norwegian).
2. Educational statistics 1962-1963 (1963) *Primary and Continuation Schools, Norway's Official Statistics A76* (Statistics Norway, Oslo), Vol 1 (in Norwegian).
3. Educational statistics 1963-1964 (1964) *Primary and Continuation Schools, Norway's Official Statistics A97* (Statistics Norway, Oslo), Vol 1 (in Norwegian).
4. Educational statistics 1964-1965 (1965) *Primary and Continuation Schools, Norway's Official Statistics A97* (Statistics Norway, Oslo), Vol 1 (in Norwegian).
5. Educational statistics 1965-1966 (1966) *Primary and Continuation Schools, Norway's Official Statistics A97* (Statistics Norway, Oslo), Vol 1 (in Norwegian).
6. Educational statistics 1966-1967 (1967) *Primary and Continuation Schools, Norway's Official Statistics A97*(1967) (Statistics Norway, Oslo), Vol 1 (in Norwegian).
7. Aakvik A, Salvanes KG, Vaage K (2010) Measuring heterogeneity in the returns to education using an education reform. *Eur Econ Rev* 54:483−500.
8. Monstad K, Salvanes KG, Propper C (2008) Education and fertility: Evidence from a natural experiment. *Scand J Econ* 110:827−853.
9. Black SE, Devereux PJ, Salvanes KG (2008) Staying in the classroom and out of the maternity ward? The effect of compulsory schooling laws on teenage births. *Econ J* 118:1025−1054.
10. Black SE, Devereux PJ, Salvanes KG (2005) Why the apple doesn't fall far: Understanding intergenerational transmission of human capital. *Am Econ Rev* 95:437−449.
11. Sundet JM, Tambs K, Magnus P, Berg K (1988) On the question of secular trends in the heritability of IQ test scores: A study of Norwegian twins. *Intelligence* 12:47−59.
12. Flynn JR (1987) Massive IQ gains in 14 nations. *Psychol Bull* 101:171−191.
13. Sundet JM, Barlaug DG, Torjussen TM (2004) The end of the Flynn effect? A study of secular trends in mean intelligence test scores of Norwegian conscripts during half a century. *Intelligence* 32:349−362.
14. Sundet JM, Tambs K, Harris JR, Magnus P, Torjussen TM (2005) Resolving the genetic and environmental sources of the correlation between height and intelligence: A study of nearly 2600 Norwegian male twin pairs. *Twin Res Hum Genet* 8:307−311.
15. Kristensen P, Bjerkedal T (2007) Explaining the relation between birth order and intelligence. *Science* 316:1717.

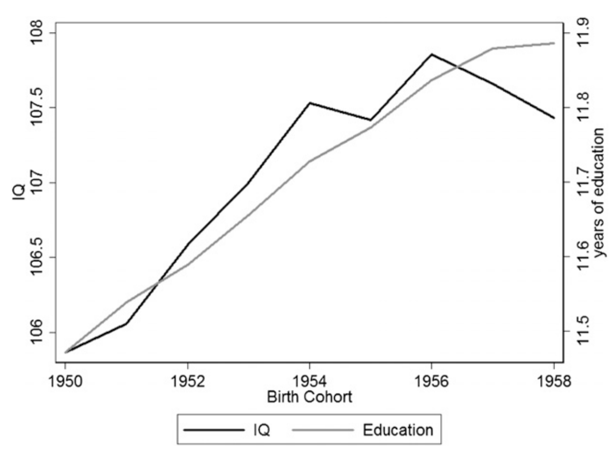

Fig. S1. Average schooling and IQ scores over birth cohorts 1950–1958.

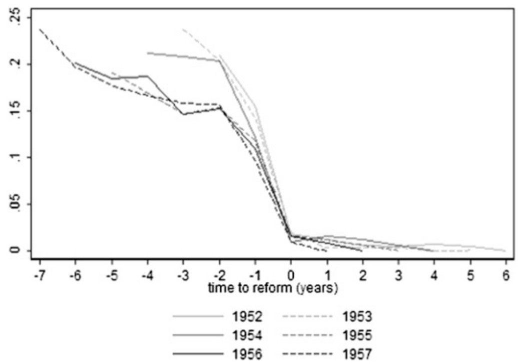

Fig. S2. Share of persons with less than 9 y of education by time to reform and reform year.

Table S1. Descriptive statistics for men born during the period 1950–1958 and sample for analysis

Birth cohort	1950	1951	1952	1953	1954	1955	1956	1957	1958	All
All men										
IQ	105.99	106.24	106.93	107.12	107.66	107.70	107.96	107.83	107.55	107.25
(SD)	(13.38)	(13.36)	(13.06)	(12.91)	(12.76)	(12.63)	(12.63)	(12.59)	(12.62)	(12.89)
Education, y	10.49	10.55	10.59	10.66	10.73	10.79	10.84	10.90	10.89	10.72
(SD)	(1.58)	(1.51)	(1.45)	(1.41)	(1.38)	(1.36)	(1.34)	(1.29)	(1.28)	(1.41)
No. observations	20,753	20,425	21,264	21,519	21,099	22,415	23,265	23,301	24,198	198,239
Analysis, %*	58.88	59.50	59.64	59.37	59.07	59.12	59.45	60.15	59.74	59.45
Sample for analysis*										
IQ	105.87	106.06	106.58	107.00	107.53	107.42	107.86	107.66	107.43	107.07
(SD)	(13.38)	(13.26)	(12.98)	(12.85)	(12.69)	(12.67)	(12.71)	(12.64)	(12.69)	(12.88)
Education, y	10.47	10.54	10.59	10.66	10.73	10.77	10.84	10.88	10.89	10.71
(SD)	(1.58)	(1.50)	(1.44)	(1.40)	(1.38)	(1.36)	(1.34)	(1.29)	(1.27)	(1.40)
No. observations	12,222	12,156	12,684	12,788	12,486	13,255	13,836	14,019	14,457	117,903
Reform, %[†]	16.37	30.52	43.18	50.71	57.55	69.48	80.7	87.48	92.82	60.15

*Men born during the period 1950–1958, for which timing of reform can be identified.
[†]Men who turned 14 y of age in a municipality with the new (postreform) schooling system.

Table S2. Effect of reform on school population aged 14 y

	Population of 14-y-olds in municipality	
	Total (*n*)	Log total (log *n*)
Reform	−5.104	−0.019
(SE)	(5.370)	(0.016)
Municipal indicators	✓	✓
Cohort indicators	✓	✓
Excluding prereform cohort	✓	✓
No. observations	107,223	107,223

Robust SEs based on the sandwich (or Huber–White) estimate of variance with clustering by municipality/cohort groupings are reported in parentheses.

Table S3. Identification of reform year

	Municipalities		Individuals	
	Frequency	Percentage	Frequency	Percentage
Unable to identify reform year because of institutional difficulties*	105	20.04	43,897	22.15
Unable to identify reform year for other reasons[†]	146	27.86	36,439	18.38
Used in analysis	273	52.10	117,903	59.47
Total	524	100.00	198,239	100.00

*Changes in municipal structure and/or inconsistencies in official statistics (details are provided in the main text).
[†]Primarily attributable to municipalities being too small to uncover clear pattern of educational attainment.

Table S4. Effect of time to reform on IQ and education

	Years of education		IQ	
	Coefficient	SE	Coefficient	SE
Years before reform				
5 or more	−0.153	0.022	−0.286	0.186
4	−0.185	0.022	−0.899	0.213
3	−0.158	0.024	−0.716	0.223
2	−0.161	0.025	−0.758	0.252
1	−0.180	0.030	−1.154	0.326
Years after reform				
1	−0.013	0.019	−0.240	0.194
2	−0.030	0.019	−0.120	0.196
3 or more	−0.026	0.022	0.197	0.244
Cohort indicators	✓		✓	
Municipal indicators	✓		✓	
No. observations	117,564		117,564	

The population for analysis consists of men born during the period 1950–1958 for whom IQ scores are available and who lived in a municipality where the timing of reform introduction could be identified. The reference for the time-to-reform variables is the first reform year (0 y after reform). Robust SEs based on the sandwich (or Huber–White) estimate of variance with clustering by municipality/cohort groupings are reported.

Table S5. Alternative estimates with last prereform cohort defined as treated

	Coefficient	SE
Reform on average IQ score	0.541	0.151
Reform on schooling	0.097	0.014
One year of schooling on IQ score	5.599	1.307
Municipal indicators	✓	
Cohort indicators	✓	
No. observations	117,564	

The timing of the reform is predated by 1 y to accommodate the partial treatment of the last prereform cohort. The population for analysis consists of men born during the period 1950–1958 for whom IQ scores are available and who lived in a municipality where the timing of reform introduction could be identified. For the quasiexperimental results, the effect of the reform on IQ score is a DID estimate and the effect of the reform on schooling and the effect of 1 y of schooling on IQ score are estimated by the 2SLS method. Under appropriate assumptions, the effect of education on IQ scores is equal to the ratio of the effect of the reform on IQ score and the effect of the reform on schooling. Robust SEs based on the sandwich (or Huber–White) estimate of variance with clustering by municipality/cohort groupings are reported.