New empirical evidence on the effect of educational tracking on social inequalities in reading achievement

European Educational Research Journal

2015, Vol. 14(3-4) 206–221

© The Author(s) 2015

Reprints and permissions:

sagepub.co.uk/journalsPermissions.nav

DOI: 10.1177/1474904115589039

eerj.sagepub.com

Jeroen Lavrijsen

HIVA – KU Leuven, Belgium

Ides Nicaise

HIVA – KU Leuven, Belgium

Abstract

One of the major imperatives behind the comprehensivisation of secondary education was the belief that postponing the age at which students are tracked in different educational routes would mitigate the effect of social background on educational outcomes. Comparative investigations of large-scale international student achievement tests in secondary education, such as PISA, have indeed suggested that individual test results depend less on social origin in countries that have postponed tracking age. However, a crucial pitfall in such cross-sectional studies is that many other factors influence the effect of social origin on achievement as well. In order to account for possible unobserved confounder bias, and to acknowledge the fact that part of the social origin effect already exists prior to the introduction of tracking, we apply a difference-in-differences analysis to data from PIRLS (primary education, 2006, $N = 33$, $n = 171.486$) and PISA (secondary education, 2012, $N = 33$, $n = 235.378$). Our results confirm that the introduction of tracking increases the effect of social origin on reading achievement between primary and secondary education. This lends further support to the argument that postponing the tracking age can foster social equity in educational achievement.

Keywords

Early tracking, social inequality, reading achievement, difference-in-differences

Corresponding author:

Jeroen Lavrijsen, HIVA – KU Leuven, Parkstraat 47 bus 5300, B-3000 Leuven, Belgium.

Email: jeroen.lavrijsen@kuleuven.be

Introduction

During the last decades, many Western countries have been addressing the comprehensivisation of their educational systems. An important ingredient was the postponement of the age at which students get tracked into different educational pathways, often a vocational and an academic track. The central argument behind this change was the belief that postponing tracking age would mitigate the strength of the association between social background and educational outcomes (Antikainen, 2006; Baldi, 2012; Henkens, 2006; Horn, 2007; Peter et al., 2010), and that comprehensivisation would thus foster the social equity of the educational system.

There are indeed good reasons to expect tracking to strengthen the link between social background and educational performance. Socially disadvantaged students seem to be disproportionately selected into less prestigious tracks, where they encounter less stimulating learning environments which may hamper their performance. The first part of the argument – children from a lower social background are more often placed in lower tracks, even after accounting for prior performance – has been empirically substantiated in studies from several countries: see Boone and Van Houtte (2012) for Flanders, Ditton and Krusken (2006) for Germany, and Duru-Bellat (2002) for France, among others. The educational ambitions of young pupils seem to be strongly influenced by the role models perceivable in their social environment, such that the aspirations of students with parents from less prestigious professions are usually more modest than those of children from high-socio-economic status parents (Breen and Goldthorpe, 1997). When tracking decisions have to be made at a young age, the parental voice is still utterly important, and the impact of socio-economic background on track placement will be strongest (Brunello and Checchi, 2007).

The second part of the argument claims that, on average, being in a lower track hampers performance, not only compared with being in an upper track, but also compared with being in an undifferentiated, heterogeneous education system (Hanushek and Woessmann, 2006). The argument here is that shifting students to a less demanding track, where the curriculum is less challenging and the learning environment far from optimal, rather leads to ignoring learning difficulties instead of adequately addressing them (Hallinan and Kubitschek, 1999; Hattie, 2002; Oakes, 1993). Moreover, educational resources tend to be unequally distributed across tracks (Darling-Hammond, 1996), and the most experienced and most capable teachers often are assigned to the high tracks, leaving the lower tracks to the less experienced teachers (OECD, 2012). Teachers in the lower tracks also tend to develop lower expectations towards their students and act accordingly (Van Houtte, 2004), for example by devoting less time to actual instruction (Hallinan, 1994a; Oakes, 1992). Likewise, the fact that students in lower tracks often end up there because of negative selection may give rise to the development of an entire class culture that gets negatively oriented towards learning, further damaging the learning climate and performance in the lower tracks (Van Houtte and Stevens, 2008, 2010). On the other hand, it has been argued that sorting pupils according to ability facilitates instruction at the right level and pace, and thus might increase the overall efficiency and performance of the educational system performance (“specialisation benefit”, for example Duflo et al., 2008; Figlio and Page, 2002; Hallinan, 1994b). This might imply that even for those in the lower tracks, tracking might be beneficial for educational performance. Similarly, the ability level of classroom peers influences individual performance as well, but neither the size nor the direction of such peer effects is yet fully clear (e.g. Dobbela et al., 2002; Hanushek et al., 2003; Hoxby, 2000).

Methodological issues in comparative research

The above arguments mostly suggest that systems with early tracking could involve less equal learning opportunities, with stronger ties between social origin and educational achievement.

Empirical comparative research indeed has seemed to corroborate this hypothesis: in their review of the literature, Van de Werfhorst and Mijs (2010) concluded that most studies found that early tracking led to stronger effects of social background on performance. The literature on which this conclusion was based has mainly exploited two different research designs. A first series of studies examined the effect of a specific educational reform in a single country. For example, Pekkarinen et al. (2009) examined the effect of a Finnish reform that postponed tracking from 12 to 16 in the 1970s. Using cognitive test scores from the military service entry exam, they showed that the reform improved the performance of those coming from socially disadvantaged backgrounds, both in terms of their absolute cognitive level and in terms of their relative position compared with those from advantaged backgrounds. The reform thus seems to have succeeded in alleviating the effect of social origin on achievement. Similar results were found in studies on educational reforms in Norway (Aakvik et al., 2010), Sweden (Meghir and Palme, 2005) and Poland (Jakubowski, 2010). By contrast, Galindo-Rueda and Vignoles (2004) argued that the demise of the British selective grammar school system led to an increase rather than a decrease in the effect of parental background on achievement, though Manning and Pischke (2006) have questioned the validity of this conclusion because of selection bias problems, as the implementation of the reform seemed to be correlated with cognitive achievement itself. An overall concern with studies evaluating a specific reform in a single country is that educational reforms usually comprise a whole bundle of policy measures, which makes it difficult to disentangle the effect of postponing tracking and to extrapolate it to other settings.

A second series of studies broadens the scope to a comparative, cross-sectional analysis of several countries, in which the social origin effects in countries with different tracking regimes are compared at a single point in time. For each country, the association between individual student achievement and social origin is determined, after which correlations between the observed associations and the features of each national educational structure can be analysed. This design has been facilitated by the proliferation of large-scale standardised student achievement datasets, in particular PISA, which deliver micro-data on achievement of 15-year-olds and their social origin for a large number of countries. The most important difficulty in these studies is that countries have many different features that could all influence the observed association between social origin and achievement, such as the level of income inequality, the degree of pre-primary school attainment or the spread of private schooling, among other things. As these could all spur the observed relationships, cross-national research has to take possible confounders accurately into account in order to provide unbiased estimates of the effect of tracking.

Several studies have used this cross-sectional approach to explain the strength of the association between social origin and achievement in terms of educational structures. For example, Duru-Bellat and Suchaut (2005) used PISA 2000 data to show that social inequalities in secondary school achievement were larger in countries that track their students at an earlier point in their careers. To accommodate for possible confounding variables, GDP/capita and the Gini-inequality index were controlled for in the models, which apparently did not change the main message. Horn (2009) came to an analogous conclusion on PISA 2003 data: the larger effect of social origin on achievement in early tracking countries survived when other relevant differences in the education system (school autonomy, use of central exams) were controlled out. Similarly, data from PISA 2009 led Bol and Van de Werfhorst (2013) to conclude that early tracking was related to less equal opportunities after controlling out differences in wealth, educational expenditure and private schooling rates. While these and other studies (Dupriez and Dumay, 2006; Dupriez et al., 2008; Schütz et al., 2008) considered cross-national differences in educational structures, Woessmann (2010) showed that also when tracking age variations between German regions were considered, early tracking seemed to be associated with a stronger impact of social origin and achievement.

However, the perennial problem with all these (and other) attempts is that no-one can ever be sure that indeed *all* relevant country-level confounders have been taken into account. Moreover, the country samples are usually rather limited in size (typically between 20 and 30 countries), which impedes a simultaneous control for many confounders in a single model. Finally, some of the possible confounders may prove difficult to measure (e.g. cultural values). This all means that cross-sectional approaches always have to acknowledge that the observed effects may be, to some extent, spurious.

Secondly, a particular concern for comparative analyses of social origin effects might be that a correlation between social origin and achievement observed in cross-sectional datasets could partly reflect a correlation between social origin and innate ability. In modern societies individuals with high (innate) ability have higher odds to reach an advantaged social status in their lives (Bell, 1976; Herrnstein, 1973). But innate ability is partly genetic and thus will be transmitted over generations (Bouchard, 2004; Neisser et al., 1996; Turkheimer et al., 2003). This means that observing a correlation between social origin and achievement does not necessarily contradict the meritocratic ideal that educational performance should only depend on ability and effort – only when social origin would have a “pure” effect on achievement (i.e. independent of ability), equality of opportunities would be violated (cf. Saunders, 1995; Breen and Goldthorpe, 1999). Hence, in simply comparing the observed social origin effect across countries, one implicitly assumes that the – unobserved – correlation between ability and social origin was the same for all countries under study. If this condition would not be met, this would bias the cross-national comparison.

The solution of difference-in-differences

We will deal with both the above concerns by combining individual data on achievement and social origin in secondary education (PISA, 15-year-olds) with similar data from a primary education assessment (PIRLS, 4th grade). This will allow us to control the association between social origin and achievement in secondary education for the association that already exists in primary education. Hence, first, such a “difference-in-differences” approach removes possible bias by unobserved confounders (e.g. income inequality), as such confounders would impact the observed association between social origin and achievement on both measurement points. By contrast, tracking age only affects the social origin effect at age 15 (as some countries have already tracked their students before this point, while others have not), but not at age 10 (as at this stage students are still untracked in all countries). Hence, by comparing the difference in the social origin effect between both measurement points, the net effect of tracking can be determined without having to include the individual confounders in the model themselves. Secondly, the difference-in-differences-approach allows us to circumvent the possible bias due to the cross-country differences in the association between innate ability and social class. Such an association will be absorbed in the social origin effect observed on both measurement points, and consequently, possible differences between countries in the size of this association will not distort estimates based on the change in the social origin effect between both points.

This approach thus bears similarities with the frequently cited article by Hanushek and Woessmann (2006), who exploited a similar methodology to show that early tracking increased the achievement gaps between low and high performers. However, Hanushek and Woessmann (2006) did not consider the effect of social origin itself, leaving this out for “further research”. To our knowledge, only two contributions (published as Working Papers) have tried to address this task. First, Ammermüller (2005) exploited data from PIRLS 2001 and PISA 2000 to confirm that tracking was associated with a larger increase in the effect of parental background between primary and secondary education. Secondly, Waldinger (2006) performed

a series of difference-in-differences analyses on data from PIRLS 2001, TIMSS 1995 and 1999, and PISA 2000 and 2003. However, Waldinger did not find a clear effect of tracking on the increase in the social origin effect, leading him to the claim that the suggestions coming from the cross-sectional studies must have been due to omitted variables bias. Both attempts, however, suffered from important weaknesses, mostly due to low country sample sizes. As only countries that participated in both the primary and the secondary assessment can be included in a difference-in-differences-model, samples turned out to be not larger than 12 countries in Ammermüller (2005) and between 8 and 14 countries (depending on the specification) in Waldinger (2006). This makes the entire set-up quite heavily dependent on the specific features of the few countries considered. For example, in Waldinger (2006) only 2–4 early tracking countries (depending on the specification) were considered. Also note that the attempts did not yet use the full richness of the social background data available; for example, Ammermüller (2005) collapsed the information on the educational level of the parents into a dichotomous variable, only indicating whether or not the parents had acquired a university degree.

Our article aims to add to the understanding of the influence of early tracking on the effect of social origin on educational achievement in the following ways. First, we will account for both the omitted variables problem and the correlation between ability and social origin, which are inherent to all cross-sectional design studies, by applying a difference-in-differences analysis, in which we correct social origin effects in secondary school for the “pre-treatment” effect already existing in primary school. Secondly, we will improve on both previous attempts that followed this route by making use of more recent student assessments: as participation in student assessments is now much more widespread, this allows us to construct much larger samples (up to 33 countries), and thus to arrive at more reliable conclusions. Finally, we will take full advantage of the richness of the social background data included in the student assessments.

Data and methodology

A difference-in-differences analysis of social origin effects necessitates data on both educational achievement and social origin at two different points in the school career, one in primary school (when all students are still educated together in every country) and one during secondary school (when students have already been tracked in some countries, but not in others). There are three large-scale student assessments that can deliver such data: for primary education we can use data from PIRLS (reading literacy) or TIMSS (science and mathematics), which are both collected in the 4th grade, while for secondary education we can use data from PISA (reading, science and mathematics), performed at 15 years of age, or again TIMSS (science and mathematics), which also collects data in the 8th grade. Regarding the latter measurement point, PISA is preferable to TIMSS as the average age of the respondents is considerably higher in PISA (15.8 years compared with 14.3 years), which means that tracking has had considerably more time to exert its influence, making the effects better observable. Regarding the primary measurement point, PIRLS is preferable to TIMSS because the definition of “achievement” is more similar between PIRLS and PISA than between TIMSS and PISA. Indeed, both PIRLS and PISA measure the proficiency to use reading skills in real-life situations (compare Mullis et al., 2006 with OECD, 2013a; also Grisay et al., 2009). By contrast, TIMSS is more focused on the extent to which a certain curriculum is mastered (Micklewright and Schnepf, 2007). From the available waves we use the data from the PIRLS 2006 wave and the PISA 2012 wave: as the time gap between both waves is close to the age difference of the respondents in their sample, both samples can be assumed to have been drawn from roughly the same population (those who were 10 in 2006 and those who were 15 in 2012 are from about the same birth cohort).¹

Table 1. Participating countries, sample sizes, and tracking ages.

Country	OECD	PIRLS sample size	PISA sample size	Tracking age	Country	OECD	PIRLS sample size	PISA sample size	Tracking age
Austria	X	5067	4755	10	Luxembourg	X	5101	5258	13
Belgium (Fl.)	X	4479	4877	12	Latvia		4162	4306	16
Belgium (Fr.)	X	4552	3720	12	Netherlands	X	4157	4460	12
Bulgaria		3864	5282	13	Norway	X	3837	4686	16
Canada	X	20,566	21,544	16	New Zealand	X	6256	4291	16
Denmark	X	4001	7481	16	Poland	X	4855	4608	16
England	X	4037	9714	16	Qatar		6681	10966	15
Spain	X	4094	25313	16	Romania		4273	5074	14
France	X	4404	4613	15	Russia		4721	5232	16
Germany	X	7899	5001	10	Scotland	X	3775	2945	16
Hong Kong		4712	4670	15	Singapore		6390	5546	12
Hungary	X	4069	4810	11	Slovenia	X	5337	5911	14
Indonesia		4774	5622	15	Slovak Republic	X	5381	4678	11
Iceland	X	3673	3509	16	Sweden	X	4394	4736	16
Israel	X	3908	5055	15	Taiwan		4590	6046	15
Italy	X	3581	31073	14	United States	X	5195	4978	16
Lithuania		4701	4618	16	TOTAL		171,486	235,378	

As Table 1 shows, a sample of 33 countries participated in both assessments, of which 23 are members of the OECD. The full dataset contains information about 170,000 respondents from PIRLS and 235,000 respondents from PISA. Table 1 also lists the tracking ages as reported by OECD (2013b).

Both PIRLS and PISA deliver individual reading achievement scores in terms of five plausible values, which we use in accordance with the recommendations from the Technical Reports, that is, all coefficients and standard errors were estimated by averaging the results from five separate regressions.

In order to determine the strength of the association between social origin and achievement, we have to make two choices. First, we have to choose a measure to represent social origin. Secondly, we have to choose a statistical measure that summarises the strength of the association between social origin and achievement.

PISA contains information on three different dimensions of social origin. First, the highest occupational status of the parents is provided as the status score that corresponds to the occupation reported by the parents (4-digit ISCO), according to the scheme of Ganzeboom et al. (1992). Second, the highest educational level of the parents is presented in terms of ISCED-level and years of education. Third, an indicator of household equipment aggregates information on the presence of 23 different items in the household (e.g. books, ICT material, newspapers). The information on these three characteristics (occupation, education and material possessions) is aggregated by means of principal component analysis and reported as the Index of Economic, Social and Cultural Status (ESCS); it is this aggregated variable that most previous research has used to represent social background, as it covers all relevant dimensions of social origin.

In PIRLS, the social background information is somewhat more limited. In contrast with PISA, PIRLS reports the occupation of the parents only on a 1-digit ISCO-scale, which is too rough for the derivation of occupational status scores. The dataset also contains information on the highest educational level of the parents in terms of the ISCED-level (which we translated into years of education with the scheme derived from PISA). There is also information on the availability of seven household items (of which three overlap with the PISA-items, including the number of books at home). The dataset also contains an assessment of the financial position of the family, as perceived by the parents. PIRLS also reports an aggregated index, aimed at combining information from several measures. However, this Index of Home Educational Resources (HER) did not use principal component analysis, but was simply constructed by placing all combinations of parental education and the number of available household items into three discrete categories. This of course leads to considerable loss of data and detail. Finally, while social background information was almost complete in the PISA dataset, some of the social background variables in PIRLS had a considerable amount of missing data (up to 20% for some variables in some countries). These missing data might be non-random, for example when socially disadvantaged parents are more likely not to answer all background questions, or when low-performing students are less attentive to fill in the entire questionnaire.

Hence, we improved the data comparability between PIRLS and PISA in two ways. First, we used imputation, in which we predict the missing values on the background information available, to reduce the number of missing values in the PIRLS dataset. We will report both imputed and not imputed data analyses. Secondly, we constructed in the PIRLS dataset a new aggregate index that resembles the PISA-ESCS-index more closely than the original (discrete) HER-index did. This new index, which we label the ESCS'-index, combines all the available information on education, financial position (as a possible proxy for occupational status) and household possessions. By using principal component analysis,² we aggregate this information in one index, which, just as the PISA ESCS-scale, aggregates all the dimensions that play a role in the effect of social origin of achievement in one continuous measure.

Finally, we will use two statistical measures to express the strength of the association between social origin and achievement. First, we use the variance in reading achievement scores that is explained by social origin, that is, the R^2 of the regression of achievement on social origin. This variable reflects how closely individual achievement is related to social origin: a large R^2 indicates that achievement is strongly predictable by social origin and, hence, that social mobility is relatively limited. Secondly, we use the slope of the regression line of achievement on social origin. This indicates how fast achievement increases when respondents come from a more advantaged social background.

Our difference-in-differences model can then be constructed as follows. First, for every country and for both assessments separately, we run a regression of individual reading achievement on social background. Second, the R^2 resp. the slope of these regressions are then used as the input for the country-level regression

$$Y_{\text{secondary},i} = a + b \cdot Y_{\text{primary},i} + c \cdot T_i + \varepsilon_i \quad (1)$$

in which $Y_{\text{secondary},i}$ is the association between social origin and achievement for PISA in country i and $Y_{\text{primary},i}$ is the association between social origin and achievement in PIRLS. T_i is our country-level tracking indicator. We use both a continuous indicator, that is, the number of years that students have spent in a tracked system before age 15, and a dichotomous indicator, that is, an 'early

Table 2. Association between social origin and reading achievement in PISA.

Country	R ²	Slope	Country	R ²	Slope
Austria	15.3	42.0	Luxembourg	15.6	37.4
Belgium – Fl.	17.6	44.2	Latvia	14.0	35.6
Belgium – Wall.	19.1	50.0	Netherlands	10.8	38.7
Bulgaria	21.9	52.5	Norway	6.3	32.7
Canada	8.1	29.7	New Zealand	16.5	51.6
Denmark	15.3	39.1	Poland	13.4	35.5
England	11.9	41.2	Qatar	5.2	28.9
Spain	12.5	31.3	Romania	16.5	38.5
France	18.7	58.2	Russia	13.1	43.1
Germany	15.0	37.5	Scotland	11.0	34.3
Hong Kong	5.2	19.9	Singapore	15.2	42.8
Hungary	20.0	42.4	Slovenia	14.2	39.6
Indonesia	6.2	17.1	Slovak Republic	24.1	55.6
Iceland	6.3	29.7	Sweden	9.1	38.1
Israel	11.2	44.3	Chinese Taipei	15.1	42.0
Italy	9.7	30.8	USA	12.6	33.3
Lithuania	11.3	31.6	Average (st. dev.)	13.3 (4.9)	38.5 (9.1)

tracking' dummy equal to 1 if and only if tracking has already taken place. We run our models both on the full sample and on the 23 OECD countries separately.

Results

As an illustration, we start by estimating the association between achievement in PISA and the aggregate variable representing social origin (ESCS-index). Table 2 presents the results for the two statistical measures of association, R^2 and slope.

Usually cross-sectional research would go on calculating the correlation between the observed measures of association and the tracking age in the countries considered. Our difference-in-differences model will instead examine how the association observed in PISA can be explained by means of the association already observed in PIRLS on one hand and tracking age on the other. Figure 1 visualises the idea (for R^2 as the measure of association): for every country, the vertical axis contains the observed association in PISA (cf. Table 2), while the horizontal axis adds the corresponding value observed in PIRLS (in which we used the original HER-index to represent social origin). Notwithstanding the differences in the variables characterising social origin between both datasets, there is a clear association between the observed values in both datasets ($\rho = 0.53, p < 0.01$). This precisely underlines the value of our approach: a large part of the social origin effects observed in secondary education is already present in primary education. However, at the same time, there is a clear difference between the regression line for the early tracking countries and the corresponding regression line for the late tracking: given the existence of a certain association in primary education, the association in secondary education is stronger if the country has adopted an early tracking system. Hence, while the social origin effect in secondary education certainly has part of its root in primary education, the introduction of early tracking does seem to increase this effect, net of the effect already existing in primary education.

We will now quantify this effect, and further verify its robustness, by running Equation (1) for different measures of association between social origin and achievement (R^2 or slope), different

Figure 1. The association between social origin and reading achievement, in primary (X-axis) and secondary (Y-axis) education.

social background variables constructed in PIRLS (the original HER-index or the newly defined ESCS'-index; for PISA we use the ESCS-index), the different ways to operationalise tracking (as a dichotomous 'early tracking' variable or as the number of years spent in a tracked system prior to the second measurement point), and different country samples (the full sample or a sample restricted to OECD countries), and to datasets with or without the use of imputation.

Table 3 starts with using the variance in achievement explained by social origin (i.e. R^2) as the measure of association, as applies this measure to the different social background variables. First, the association in PISA is indeed significantly related to the size of the association already observed in PIRLS, which again underlines the value of a difference-in-differences model. Second, although their significance varies somewhat depending on the specification, all estimates of the effect of tracking are consistently pointing in the same direction: early tracking always leads to a higher share of the variance in reading achievement that can be explained by social origin. This effect is observed in all specifications, regardless of which variable we chose to measure social background in PIRLS (HER or ESCS'), regardless of the way we characterise tracking (dichotomous or continuous) and regardless of the use of imputation (though imputation increases the fit of the model). Over all specifications, early tracking leads to an increase between 3.1 and 4.6 percentage points in the size of the variance explained by social origin, which is both statistically significant and practically relevant (e.g. close to one standard deviation in the full sample, see Table 2). Equivalently, every year a student spent in a tracked system leads to an about 1 percentage point higher explained variance in secondary school achievement. Finally, note that our two improvements to the social background data in PIRLS indeed seem to increase the statistical power of our models: both the use of the ESCS'-scale, which aggregates all social background data into one continuous measure instead of the discrete HER-scale, and the use of data imputation increases the R^2 of our difference-in-differences-model, leading to more accurate predictions. Hence, in what follows we will use the ESCS' as the preferred measure of social origin in PIRLS and perform our analyses on the imputed dataset.

Table 3. The influence of tracking on the increase in the social origin effect (R^2) between primary and secondary education.

	Social origin in PIRLS measured by									
	HER-index			ESCS ¹ -index			ESCS ¹ -index, with imputation			
	All	OECD	All	OECD	All	OECD	All	OECD	All	OECD
Intercept	6.63	7.01	6.79	7.00	6.19	6.90	6.16	6.70	5.95	5.88
Social origin effect in PIRLS	0.65***	0.60**	0.65***	0.62**	0.47***	0.41**	0.49***	0.44**	0.49***	0.51***
Tracking measure										
Early tracking	4.63***	4.38**			3.93***	3.83**			3.48**	3.14*
Number of years tracked			1.12***	1.08**			0.90**	0.92*		0.75*
R^2 (adj.)	0.50	0.43	0.44	0.40	0.58	0.50	0.53	0.48	0.61	0.57
N	32	22	32	22	31	21	31	21	31	21

Note: Dependent variable: social origin effect in PISA. Measure of association between social origin and achievement: R^2 . *** $p < 0.01$; ** $p < 0.05$; * $p < 0.1$.

Table 4. The influence of tracking on the increase in the social origin effect (slope) between primary and secondary education.

	Social origin in PIRLS measured by ESCS'-index, with imputation			
	All	OECD	All	OECD
Intercept	20.50	17.95	20.83	18.17
Social origin effect in PIRLS	0.56***	0.64**	0.57***	0.64**
Tracking measure				
Early tracking	7.16***	6.05**		
Number of years tracked			1.54**	1.42*
R ² (adj.)	0.46	0.39	0.39	0.35
N	31	21	31	21

Note: Dependent variable: social origin effect in PISA. Measure of association between social origin and achievement: regression line slope. *** $p < 0.01$; ** $p < 0.05$; * $p < 0.1$.

Table 4 shows that we get very similar results when we use the regression slope, instead of the explained variance (R^2), as the measure of association between social origin and achievement. Again, social origin already has a large effect on achievement in primary education and this effect persists in secondary education. But yet again, on top of this pre-existing social inequality, the presence of early tracking leads to sizeably steeper gradients (the difference of about 7 points is again both statistically significant and practically relevant, cf. close to one standard deviation in the full sample, Table 2). Note that the model fit (adjusted R^2) of the difference-in-differences-models using regression slopes (Table 4) are somewhat lower than those for the corresponding models that used explained variance as the measure of association (i.e. the last four columns from Table 3).

However, the fact that early tracking increases the slope of the regression line does not necessarily imply that it would hinder the educational achievement of socially disadvantaged students. For example, even when tracking would be beneficial to every student (cf. the alleged “specialisation benefit”), it could still make the regression line steeper as long as the benefit is largest for socially advantaged students. Hence, we additionally perform an analysis of the performance growth for different social groups of students. We identify four groups of students on the basis of their place in the social status distribution: those coming from the 10% with respect to the 25% most disadvantaged families on one side, and those coming from the 10% with respect to the 25% most advantaged on the other. We now compare the absolute achievement level of those groups in PISA with the results for the corresponding social group in PIRLS. Table 5 confirms that early tracking seems indeed detrimental to the fate of socially disadvantaged students. The socially disadvantaged students score between 16.0 (25th percentile) and 20.9 (10th percentile) points worse on the PISA scale if they have been subjected to an early tracking regime, after adjusting for performance in primary education. This clearly is a sizeable loss, as in PISA the effect of 1 year of schooling is estimated to be around 40 points (OECD, 2013b). At the other end of spectrum, advantaged students seem to gain only very little (only a few points, indiscernible from zero) from early tracking. Hence, the observation that tracking boosts the social gradient seems to be mainly due to the loss that is suffered by disadvantaged students.

Finally, we check whether similar relationships can be observed when we quantify social background by two of its ingredients separately (the educational level of the parents and the number of books at home). While this approach has the obvious disadvantage that it uses only part of the available information, it has the advantage that the separate characteristics (which have both been shown to be important aspects of social background) are defined completely comparably across

Table 5. Effect of tracking on the absolute achievement of different groups in the social status distribution.

	10% most disadvantaged students	25% most disadvantaged students	25% most advantaged students	10% most advantaged students
Intercept	130.86	136.72	193.33	200.68
Achievement score in PIRLS	0.63***	0.63***	0.61***	0.61***
Early tracking	-20.86*	-16.07 ($p = 0.14$)	3.03 ($p = 0.76$)	2.49 ($p = 0.81$)
R^2 (adj.)	0.39	0.46	0.59	0.56
N	31	31	31	31

Note: Dependent variable: reading achievement score in PISA. Social origin in PISA measured by ESCS, in PIRLS by ESCS', with imputation, full sample. *** $p < 0.01$; ** $p < 0.05$; * $p < 0.1$.

Table 6. Effect of social origin quantified by one of its dimensions only.

	Social origin measured by educational level	Social origin measured by number of books at home
Intercept	3.44	5.99
Social origin effect in PIRLS	0.18*	1.03***
Early tracking	2.12**	1.72
R^2 (adj.)	0.28	0.67
N	32	33

Note: Dependent variable: social origin effect in PISA. Measure of association: R^2 , with imputation, full sample. *** $p < 0.01$; ** $p < 0.05$; * $p < 0.1$.

both datasets. Table 6 shows that even when social origin is quantified by just one dimension, early tracking consistently leads to a stronger link between social background and achievement, although the effect is now somewhat less pronounced than before (possibly due to a loss of detail).

Discussion and conclusion

In this paper, we studied how postponing the age of tracking in some countries may have succeeded in reducing the strength of the association between social background and achievement. In order to account for unobserved confounder bias and to acknowledge the fact that part of the social origin effect exists already in primary education, we applied a difference-in-differences analysis to social origin and reading achievement data from PIRLS 2006 (primary education) and PISA 2012 (secondary education). Our results indicate that countries that pupils tracked at an early age undergo stronger effects of social origin on individual achievement in secondary school, net of differences existing before the tracking age. This pattern is consistently reproduced over all specifications, exploiting different measures of association, different social background variables and different operationalisations of the predictor variable. In particular, we observe that early tracking is detrimental to the educational opportunities of socially disadvantaged students, while it does not seem to affect the achievement of their more advantaged peers. These findings thus confirm, from a new methodological perspective, earlier findings on the negative impact of early tracking on equality of opportunities (see Bol and Van de Werfhorst (2013) for a review).

What do these findings teach us about the desirability of a comprehensivisation of secondary education? Here, we would want to make three points. The first is that a large-scale international comparison such as ours has to rely on a relatively crude categorisation of nation-specific educational practices. While we defined tracking in terms of the age of first selection, national practices are often more subtle than such a quantification suggests. For example, Dupriez et al. (2008) demonstrated that the different late-tracking countries have confronted the challenge of the resulting heterogeneous classrooms in different ways: while, for example, the Scandinavian countries adopted forms of individualised teaching (differentiated teaching, tutoring) to promote the learning of all students, other countries (e.g. France) ‘solved’ the challenge mostly through an increased use of grade retention. As the latter has its own undesirable consequences, a mere delay of the tracking age does not suffice to arrive at more equitable outcomes (see Figure 1), but should be accompanied with efforts to raise the educational performance of disadvantaged students.

Similarly, even within a system of early selection the adverse effects of tracking can probably be mitigated to some extent. First, tracking can be made less ‘rigid’ by facilitating promotion to a ‘stronger’ track at a later stage in the educational career. Second, the relation between track placement and overall cognitive development can be limited by leaving classrooms untracked for certain specific subjects (e.g. history or social science, see Van de Werfhorst, 2014). Third, it has been argued that reliance on central examinations for track placement (instead of mere parental preference) and school quality control (making schools more eager to invest in lower-track students) reduces the effect of social background on academic achievement in tracked systems (Bol et al., 2014).

Finally, note that the design of the educational system itself can probably not be seen as fully independent from its broader socio-economic context (Dupriez and Dumay, 2006). In recent decades many ties between the design of the educational system and that of the broader welfare state (Esping-Andersen, 1990) or economic system (Estevez-Abe, 2001) have been detected (Allmendinger and Leibfried, 2003; Andres and Pechar, 2013; Hega and Hokenmaier, 2002; Peter et al., 2010; West and Nikolai, 2013). If educational systems, welfare provision and labour market layout form more or less coherent “regimes”, the logic of path-dependence might explain why educational systems confronted with common challenges have opted for their own specific answers (Green et al., 1999).

However, even with this caution in mind, the present analysis, together with the earlier evidence, underlines the fact that educational structures, and in particular the design of secondary education, are important to understand the issue of equality of opportunity. In this sense, postponing the introduction of rigid tracking seems to remain a major imperative for educational system reform today.

Declaration of conflicting interest

The authors declare that there is no conflict of interest.

Funding

This research was funded by the Flemish Government (Policy Research Centre ‘Educational and School Careers’).

Notes

1. Note that the average coverage rate (i.e. the percentage of the population that is included in the sample design) is somewhat lower in PISA than in PIRLS: in some countries, a substantial proportion of the 15-year-olds has already left the education system. Very low retention rates might bias the estimation of social origin effects, as they would leave the most disadvantaged youngsters out of the sample. This could mainly be an issue for developing countries in which education at the age of 15 is not yet universal. In our sample, one country (Indonesia) indeed has a low retention rate for PISA (63%). Deleting this

country does not, however, change our results (e.g. the estimate for the early tracking effect in Table 3 (with social origin in PIRLS measured by HER) would be 4.53*** instead of 4.63***).

2. The coefficients of the first principal component of the principal components analysis for the nine underlying social background variables were:

- Financial position of the family: 0.20
- Years of education (maximum of both parents): 0.46
- Number of books at home: 0.44
- Availability of PC: 0.44
- Availability of own study desk: 0.34
- Availability of own books: 0.25
- Availability of newspaper: 0.18
- Availability of own room: 0.25
- Availability of own mobile phone: 0.31

References

- Aakvik A, Salvenes KG and Vaage K (2010) Measuring heterogeneity in the returns to education using an education reform. *European Economic Review* 54: 483–500.
- Allmendinger J and Leibfried S (2003) Education and the welfare state: The four worlds of competence production. *Journal of European Social Policy* 13: 63–81.
- Ammermüller A (2005) *Educational Opportunities and the Role of Institutions*. ZEW Discussion Papers 05–44.
- Andres L and Pechar H (2013) Participation patterns in higher education: A comparative welfare and production regime perspective. *European Journal of Education* 48: 247–261.
- Antikainen A (2006) In search of the Nordic model in education. *Scandinavian Journal of Educational Research* 50: 229–243.
- Baldi G (2012) Schools with a difference: Policy discourses and education reform in Britain and Germany. *West European Politics* 35: 999–1023.
- Bell D (1976) The coming of the post-industrial society. *The Educational Forum* 40: 574–579.
- Bol T and Van de Werfhorst H (2013) Educational systems and the trade-off between labor market allocation and equality of educational opportunity. *Comparative Education Review* 57: 285–308.
- Bol T, Witschge J, Van de Werfhorst H, et al. (2014) Curricular tracking and central examinations: Counterbalancing the impact of social background on student achievement in 36 countries. *Social Forces* 93: 1–28.
- Boone S and Van Houtte M (2012) Social inequalities in educational choice at the transition from primary to secondary education: A matter of rational calculation?. *Kultura i Edukacja / Culture and Education* 91: 188–214.
- Bouchard TJ (2004) Genetic influence on human psychological traits. A survey. *Current Directions in Psychological Science* 13: 148–151.
- Breen R and Goldthorpe JH (1997) Explaining educational differentials towards a formal rational action theory. *Rationality and Society* 9: 275–305.
- Breen R and Goldthorpe J (1999) Class inequality and meritocracy: A critique of Saunders and an alternative analysis. *The British Journal of Sociology* 50: 1–27.
- Brunello G and Checchi D (2007) Does school tracking affect equality of opportunity? New international evidence. *Economic Policy* 22: 781–861.
- Darling-Hammond L (1996) The right to learn and the advancement of teaching: Research, policy, and practice for democratic education. *Educational Researcher* 25: 5–17.
- Ditton H and Krusken J (2006) Der Übergang von der Grundschule in die Sekundarstufe. *Zeitschrift für Erziehungswissenschaft* 9: 348–372.
- Dobbelsteen S, Levin J and Oosterbeek H (2002) The causal effect of class size on scholastic achievement: Distinguishing the pure class size effect from the effect of changes in class composition. *Oxford Bulletin of Economics and Statistics* 64: 17–38.

- Duflo E, Dupas P and Kremer M (2008) *Peer effects, teacher incentives, and the impact of tracking: Evidence from a randomized evaluation in Kenya*. National Bureau of Economic Research.
- Dupriez V and Dumay X (2006) Inequalities in school systems: effect of school structure or of society structure?. *Comparative Education* 42: 243–260.
- Dupriez V, Dumay X and Vause A (2008) How do school systems manage pupils' heterogeneity? *Comparative Education Review* 52: 245–273.
- Duru-Bellat M (2002) *Les inégalités sociales à l'école: genèse et mythes*. Paris: Ed. PUF.
- Duru-Bellat M and Suchaut B (2005) Organisation and context, efficiency and equity of educational systems: What PISA tells us. *European Educational Research Journal* 4(3): 181–194.
- Esping-Andersen G (1990) *The Three Worlds of Welfare Capitalism*. Cambridge: Polity Press
- Estevez-Abe M (2001) Social protection and the formation of skills: A reinterpretation of the welfare state. In: Hall PA and Soskice D (eds) *Varieties of Capitalism. The Institutional Foundations of Comparative Advantage*. Oxford: Oxford University Press, pp. 145–183.
- Figlio DN and Page ME (2002) School choice and the distributional effects of ability tracking: Does separation increase inequality?. *Journal of Urban Economics* 51: 497–514.
- Galindo-Rueda F and Vignoles A (2004) *The heterogeneous effect of selection in secondary schools: Understanding the changing role of ability*. IZA Discussion paper series.
- Ganzeboom HB, De Graaf M and Treiman DJ (1992) A standard international socio-economic index of occupational status. *Social Science Research* 21: 1–56.
- Green A, Leney T and Wolf A (1999) Convergences and divergences in European education and training systems. Research Project commissioned by the European Commission Directorate-General XXII.
- Grisay A, Gonzalez E and Monseur C (2009) Equivalence of item difficulties across national versions of the PIRLS and PISA reading assessments. *IERI Monograph Series: Issues and Methodologies in Large-Scale Assessments* 2: 63–84.
- Hallinan MT (1994a) School differences in tracking effects on achievement. *Social Forces* 72: 799–820.
- Hallinan MT (1994b) Tracking: From theory to practice. *Sociology of Education* 67: 79–84.
- Hallinan MT and Kubitschek WN (1999) Curriculum differentiation and high school achievement. *Social Psychology of Education* 3: 41–62.
- Hanushek EA, Kain JF, Markman JM, et al. (2003) Does peer ability affect student achievement?. *Journal of Applied Econometrics* 18: 527–544.
- Hanushek EA and Woessmann L (2006) Does educational tracking affect performance and inequality? Differences-in-differences evidence across countries. *Economic Journal* 116: C63–C76.
- Hattie JA (2002) Classroom composition and peer effects. *International Journal of Educational Research* 37: 449–481.
- Hega GM and Hokenmaier KG (2002) The welfare state and education: A comparison of social and educational policy in Advanced industrial societies. *German Policy Studies* 2: 143–173.
- Henkens B (2006) *Steekspel tussen traditie en vernieuwing. Dertig jaar polemie over de structuur van het secundair onderwijs (1963-1992)*. PhD thesis, KU Leuven.
- Herrnstein R (1973) *IQ in the Meritocracy*. Boston: Little, Brown.
- Horn D (2007) A hypothesis on how educational regimes differ. Unpublished paper.
- Horn D (2009) Age of selection counts: A cross-country analysis of educational institutions. *Educational Research and Evaluation* 15: 343–366.
- Hoxby C (2000) *Peer effects in the classroom: Learning from gender and race variation*. National Bureau of Economic Research.
- Jakubowski M (2010) *The Impact of the 1999 Education Reform in Poland*. World Bank.
- Manning A and Pischke JS (2006) *Comprehensive versus Selective Schooling in England and Wales: What do we know?* CEE DP 66.
- Meghir C and Palme M (2005) Educational reform, ability, and family background. *The American Economic Review* 95: 414–424.
- Micklewright J and Schnepf SV (2007) *Inequality of Learning in Industrialised Countries*. IZA Discussion Paper No. 2517, Bonn, December.

- Mullis I, Kennedy A, Martin M, et al. (2006) *Assessment Framework and Specifications PIRLS 2006*. Chestnut Hill: TIMSS & PIRLS International Study Center.
- Neisser U, Boodoo G, Bouchard Jr, TJ, et al. (1996) Intelligence: Knowns and unknowns. *American Psychologist* 51: 77–101.
- Oakes J (1992) Can tracking research inform practice? Technical, normative, and political considerations. *Educational Researcher* 21: 12–21.
- Oakes J (1993) Creating middle schools: Technical, normative, and political considerations. *Elementary School Journal* 93: 461–480.
- OECD (2012) *Equity and Quality in Education*. Paris: OECD.
- OECD (2013a) *PISA 2012 Assessment and Analytical Framework: Mathematics, Reading, Science, Problem Solving and Financial Literacy*. Paris: OECD.
- OECD (2013b) *PISA 2012 Results: What Makes Schools Successful?*. Paris: OECD.
- Pekkarinen T, Uusitalo R and Kerr S (2009) *School Tracking and Development of Cognitive Skills*. IZA Discussion Paper No. 4058, Bonn, March.
- Peter T, Edgerton JD and Roberts LW (2010) Welfare regimes and educational inequality: A cross-national exploration. *International Studies in Sociology of Education* 20: 241–264.
- Saunders P (1995) Might Britain be a meritocracy? *Sociology* 29: 23–41.
- Schütz G, Ursprung HW and Woessmann L (2008) Education policy and equality of opportunity. *Kyklos* 61: 279–308.
- Turkheimer E, Haley A, Waldron M, et al. (2003) Socioeconomic status modifies heritability of IQ in young children. *Psychological Science* 14: 623–628.
- Van de Werfhorst H (2014) Changing societies and four tasks of schooling: Challenges for strongly differentiated educational systems. *International Review of Education* 60: 123–144.
- Van de Werfhorst H and Mijs JJ (2010) Achievement inequality and the institutional structure of educational systems: A comparative perspective. *Annual Review of Sociology* 36: 407–428.
- Van Houtte M (2004) Tracking effects on school achievement: A quantitative explanation in terms of the academic culture of school staff. *American Journal of Education* 110: 354–388.
- Van Houtte M and Stevens PA (2008) Sense of futility the missing link between track position and self-reported school misconduct. *Youth and Society* 40: 245–264.
- Van Houtte M and Stevens PA (2010) The culture of futility and its impact on study culture in technical/vocational schools in Belgium. *Oxford Review of Education* 36: 23–43.
- Waldinger F (2006) *Does tracking affect the importance of family background on students test scores?*. Working Paper, LSE.
- West A and Nikolai R (2013) Welfare regimes and education regimes: Equality of opportunity and expenditure in the EU (and US). *Journal of Social Policy* 42: 469–493.
- Woessmann L (2010) Institutional determinants of school efficiency and equity: German states as a microcosm for OECD countries. *Journal of Economics and Statistics* 230: 234–270.

Author biographies

Jeroen Lavrijsen is a senior research associate at HIVA – KU Leuven (Research Institute for Work and Society). He investigates the effect of educational systems on educational outcomes, with special attention to patterns of social inequality.

Ides Nicaise is a senior lecturer in ‘Education and Society’ issues at the Faculty of Psychology and Education Sciences (KU Leuven). He also works as a research manager at HIVA and leads the (inter-university) Policy Research Centre of Educational and School Careers. His research focuses mainly on the economics of education as well as on poverty and social exclusion.