Amsterdam Centre for Inequality Studies (AMCIS)
Working Paper Series

Curricular Tracking and Central Examinations: Counterbalancing the Impact of Social Background on Student Achievement in 36 Countries*

Thijs Bol
Jacqueline Witschge
Herman G. Van de Werfhorst
Jaap Dronkers

AMCIS Working Paper 2013/1

February 2013

* Direct correspondence to Thijs Bol (t.bol@uva.nl) at University of Amsterdam, Oudezijds Achterburgwal 185, 1012 DK, Amsterdam, The Netherlands.

Abstract

Tracked educational systems are associated with a greater social inequality in children's educational achievement. Until now research has assumed that the impact of tracking on the inequality of educational opportunity is independent of other educational institutional features. Using data from the 2006 PISA survey, we study how central examinations affect the association between tracking and inequality. We find that parent's social class has a larger effect on student achievement in systems without central examinations, whereas in systems with central examinations this relationship is attenuated. We argue that central examinations help hold schools accountable for their performance, thereby making it more likely for schools to allocate students to tracks and reward them on the basis of objective indicators, thereby reducing the impact of parental status on children's performance.

INTRODUCTION

There has been much comparative research which has shown that the level of inequality in educational outcomes differs across Western societies. As well as exploring the effects of socio-economic differences within individual countries, researchers have shown that the varying levels of educational inequality in different countries are affected by the organization of education within those different countries (Allmendinger 1989; Shavit and Müller 1998; Kerckhoff, 2001). One aspect of educational systems that seems to have a particular effect on students' achievement is curricular tracking, namely the way in which students are, or are not, 'sorted' into different types of education or school tracks (Marks 2005; Brunello and Checchi 2007; Horn 2009; Van de Werfhorst and Mijs 2010). Educational systems which allocate students to a large number of different educational tracks at an early age are said to increase social inequalities, mostly because such 'tracking' magnifies the impact of socio-economic status on educational achievement and attainment (Marks, 2005; Horn, 2009). These results are backed up by various studies which have shown such inequality has been reduced in countries which moved from tracked to comprehensive education systems (Gamoran 1996; Gamoran and Weinstein 1998; Duru-Bellat and Kieffer 2000; Meghir and Palme 2005; Pekkarinen, Uusitalo and Kerr 2009).

However, studies dealing with the effect of curricular tracking on educational inequality have generally overlooked additional institutional characteristics that also influence the effect of social origin on educational achievements. Such studies have predominantly assumed that curricular tracking alone is responsible for magnifying the effect of social inequality on educational achievement, and that other institutional characteristics are irrelevant in explaining how tracking relates to social inequality (but see Marks 2005; Dunne, 2010). With this research, we aim to fill this gap, and to that end we argue that at least one other aspect of cross-national institutional variation is also crucial for how socio-economic background affects educational achievement, namely the extent to which educational systems are nationally standardized (Erikson and Jonsson 1996; Wössmann 2003).

An important aspect of standardization is central exit examination. Research has already shown that central examinations stimulate schools to optimize their performance by offering incentives to perform and a more objective signaling of academic achievement (Bishop 1997; Fuchs and Wössmann 2007; Horn 2009). This may be particularly helpful for students with a lower socio-economic status, who tend to depend more on the educational system for their learning than students from more advantageous social backgrounds (Coleman 1966; Van de Werfhorst and Mijs 2010).

In this article we argue that central exams not only serve to reduce social inequality in school achievement, but also mitigate the effect of tracking on inequality. Some evidence for this claim can be found in a comparative study on the United States and Israel by Ayalon and Gamoran (2000), which argues that inequality in educational achievement by social class is more strongly affected by tracking in an unstandardized educational system without central examinations (the United States) than in a standardized system with high-stakes central exams (Israel). However, no study to date has empirically tested this hypothesis for a larger number of countries.

The main objective of this study is therefore to examine the effect of tracking on the inequality of educational opportunity in the context both of countries which have central exams and those which do not. The study will focus on secondary education as it is here that the institutionalization of curricular tracking has progressed most strongly.[1] Our work starts from the widely supported assumption that student performance is at least partly affected by the organization of educational systems, and will consider how students' achievement is impacted by curricular tracking between schools and central exit examinations, both independently and in interaction with each other.

In order to find an answer to this question, we will perform a multilevel analysis of data from OECD’s ‘Programme for International Student Assessment’ (PISA) of 2006. This dataset allows for sufficient levels of between-country variation in terms of tracking and central exit examinations. We examine data at two (students in countries) and at three levels (students in schools in countries). This combination of two-level and three-level models allows us to consider whether in more strongly-tracked systems, greater inequalities are found between families within a school, or between different schools. We will argue that the inclusion of the school-level is desirable in order to get more correct estimators of tracking and central examination, because these system characteristics are mediated by allocation and teaching/learning processes of students into schools. However, we start with a two-level analysis because that is still the state of the art and to demonstrate the difference.

CURRICULAR TRACKING AND INEQUALITY IN LEARNING

An increasing body of internationally standardized datasets has allowed researchers to compare educational outcomes across countries. Empirical findings have shown that alongside the effects of individuals and schools, educational systems do indeed contribute to inequality of educational achievement (for an overview, see Van de Werfhorst and Mijs 2010). Within the different education systems, curricular tracking— also referred to as 'ability grouping', 'streaming', 'sorting', 'differentiation', and 'placement policies'—is one of most often studied features of educational systems.

Tracking has been conceptualized as practices of allocating students into school programs or classes which are more homogeneous in terms of 'cognitive ability' (Oakes 2005; Marks 2006; Horn 2009). The practices occur in a variety of ways: students can be allocated to separate schools each offering a different curriculum, students can be divided between different types of education (or “curricular tracks”) within the same school—often labeled academic, vocational, or general—or students can be put into classes at different levels (or “streams”) for the same subject within the same school. Up to now, cross-national research has mostly focused on tracking featuring different types of education (either between or within schools), as this is a characteristic that varies more between countries than within-school ability grouping (Van de Werfhorst and Mijs 2010). In the present work, we argue that tracking is not a binary variable but rather gradated, meaning that educational systems can be more or less tracked. The level of tracking depends on several factors, all of which are important. An educational system with a 'high' level of formalized tracking would thus feature earlier selection, a tracked curriculum which lasts longer, and a greater number of tracks.[2]

As mentioned above, tracking occurs not only between but also within school types. Within-school type tracking can however hardly be seen as a variable at the level of *national* systems, as in educational systems which have within-school tracking, for example the United States, tracking policies actually differ widely between schools

(Oakes 2005). Moreover, although it could be argued that all educational systems inherently have some form of tracking, the phenomenon is arguably most strongly institutionalized in educational systems which feature curricular tracking between schools, as in these societies there are stronger beliefs that early allocation of students to a large number of tracks is acceptable (LeTendre, Hofer and Shimizu 2003). For this study, we see an educational system in which students of different academic abilities are placed in separate schools, for all subjects, and for multiple years, as more strongly tracked than a system where tracks exist within schools, are subject-specific, or can be changed between school grades. Curricular tracking can be more or less institutionalized, and personal beliefs about the legitimacy of tracking are strongly related to the extent to which tracking is organized following formal national-level institutions. In strongly-tracked systems such as Germany or the Netherlands, it would be impossible for a student not to know which track they are in, whereas in other systems, for example the American tracking system, it has indeed been reported that some students do not know which track they are in, or are not even aware of the system (Oakes 2005). We follow the approach of earlier comparative studies, and focus exclusively on the more overt forms of curricular tracking, that is, curricular tracking between school types.

One important motivation which has been cited for increasing the level of tracking is that this allows for a heterogeneous student population to choose the education that best matches their interest and learning ability thereby maximizing the average student achievement (Hanushek and Wössmann 2006). However, this claim is not well substantiated empirically (e.g., Schütz et al. 2008), and many studies instead point to the detrimental effects of tracking on the performance of less advantaged students, thereby exacerbating social inequalities (Hallinan 1994; Van de Werfhorst and Mijs 2010). Tracking has two key effects on the relation between socio-economic background and student achievement.

First, the 'social selection effect', meaning that socio-economic background plays a more significant role in the allocation of students to different tracks, especially when this occurs at a younger age. This ‘selection effect’ of curricular tracking has been explained by the life course hypothesis, which states that one’s social background has a greater effect on ones success at a younger age, and declines as one becomes older (Shavit and Blossfeld 1993). Indeed studies have found that the influence of social background on the choice of school type is greater when educational systems are tracked relative to comprehensive systems, as students are placed in one of a large number of hierarchical tracks at a young age (Marks 2005; Hanushek and Wössmann, 2006).[3]

Second, factors including socio-psychological aspects, means of instruction, resources available, and peer effects have been shown to play a role here. For example, students who start out in more demanding curricular tracks tend to show greater gains in learning (Marks 2006). This is the track effect that is relevant to students' success after selection into tracks has occurred. One important reason why different curricular tracks lead to varying levels of student achievement is the factor of 'resource endowment' which has strong implications for student performance (Brunello and Checchi 2007). In terms of achievement, higher curricular tracks tend to benefit from better educational resources (Figlio and Page 2002). The impact of social selection and the effects of tracking as we set out above thus rightly suggest that the more tracked an educational system is, the more student performance is affected by social background (*hypothesis 1a*).

In addition to the question of social selection, when studying the relation between educational systems and the inequality of opportunity it is also important to consider school context as well. It is possible that classroom composition and peer interactions might have different effects in tracked and comprehensive schools for students from different social backgrounds (Richer 1976; Minter Hoxby 2000; Wilkinson et al. 2002; Entorf and Lauk 2006). For students educated in schools from different tracks, school environments are a key source of variation in achievement levels. Recently Dronkers, Van der Velden, and Dunne (2012) showed that school characteristics affect the relation between socio-economic background and achievement. Moreover, studies have shown that school characteristics seem to mitigate some of the negative effects of educational system characteristics as outlined earlier. Simply put, one of the reasons why tracked education systems show a stronger correlation between socio-economic background and achievement is that school effects are more important in such systems.

However, contrary to what one might expect from the evidence discussed above, in tracked systems inequality of educational achievement by social background at the individual level did not appear to be higher, and students from a higher socio-economic background actually achieve *less* in highly tracked educational systems than their counterparts in comprehensive systems (Dunne 2010). Although highly tracked systems offer less equal educational opportunity this effect is moderated by the way in which the student bodies of individual schools are composed. These school-composition effects are more significant in tracked educational systems, where students are divided between curricular tracks. In contrast, the within-school effect of socio-economic background on an individual's educational performance is therefore found to be stronger in more comprehensive systems (Dronkers, Van der Velden, and Dunne 2012).

In this article we will not focus on the role played by individual schools in either providing opportunities for students, or placing constraints on their educational performance. However, in the analyses we will consider the school-level in order to correctly estimate the effects of tracking and central examination. One key reason that nationwide curricular tracking causes inequality of educational opportunity is the allocation of students in different curricular tracks, and therefore when we also consider the school-level, our hypothesis changes. Following the work of Dunne (2010) and Dronkers, Van der Velden, and Dunne (2012), if we take the school-level into account, we expect that socio-economic background will have less of an effect on student performance in educational systems with more tracking (*hypothesis 1b*).

HOW CENTRAL EXAMINATIONS INTERFERE

In the present work, we are mainly interested in the question of whether the effect of tracking on inequality of educational opportunity changes significantly depending on the level of standardization of educational systems, and more specifically how this is affected by the existence of central exit examinations.

Central examinations

Standardization can be defined as “the degree to which the quality of education meets the same standards nationwide” (Allmendinger 1989: 233). An important aspect of standardization is school accountability, or the degree to which schools have standardized incentives to perform (Wössmann 2003). Such incentives to make the performance of schools more accountable include the monitoring of teacher quality, inspections, and nationally recognized exams (most notably school-leaving exams). Of these, central examinations are one of the most vital tools for accountability, as they implement nationwide transparency in performance (Fuchs and Wössmann 2007). For the present article, we will use Bishop's (1997) definition of central exams as "curriculum-based external exit exams" (CBEEEs). A more detailed description of CBEEEs is given in the "Variables" section of this article.

It is argued that central exams impact student performance by improving the 'signaling' of academic achievement (Bishop 1997). If students take central exams, this allows one to compare performance—both between individual students and between schools.[4] One main reason for this is that central examinations force schools themselves to optimize their students' performance. This is because the results of standardized examinations are reported and known to parents, and therefore competition between schools increases and there is more incentive to perform(Bishop 1997).[5] Chiang (2009) showed that school accountability improved student performance through an improved allocation of school resources.[6]

The interaction between tracking and central examinations

How do we expect central examinations to alter the "tracking effect?" So far only one study has looked into this possible interaction. In their study, Ayalon and Gamoran (2000) compared Israel and the United States, and found that in the United States, curriculum tracking led to lower overall achievement and more inequality of performance, whereas the effects were the opposite for Israel. The study suggested that the difference was because "Israel's national examinations create incentives for achievement among teachers and students in all levels of academic courses," whereas the "absence of incentives for hard work outside of the highest-level classes" in the United States leads to lower performance of students in the lower tracks (Ayalon and Gamoran 2000: 55).

In this article, we take up their suggested explanation, and consider the different impact of tracking under different levels of standardization (i.e. the presence or absence of central exams) for a large number of countries. In countries with central exit examinations, academic performance becomes relatively more important, as schools are held more directly accountable for their performance. In these systems it is in the interest of schools to have students graduate at the highest possible level of achievement, and students are therefore allocated to educational programs more according to academic criteria. In the absence of central examinations, by contrast, academic achievement is a less important factor in allocating students to tracks, and the secondary effects of parental background instead play a more important role.

It is also possible that considerations of "status maintenance," the idea that children tend to choose educational options that minimize the risk of downward mobility (Breen and Goldthorpe 1997), may outweigh students' level of academic ability as a factor in choosing a track. Middle-class children could then enter a more academic track even if their ability is not quite sufficient. For intermediate ability levels in particular this will lead to substantial differences depending on children's social classes (Boudon 1974; Breen and Yaish 2006). However, when schools are held accountable by means of central examinations leading to more weight being placed on students' ability levels. Our argument here thus is not that selection on scholastic achievement makes track placement fully equal across social classes, as social classes differ in their performance. Rather, inasmuch as social class differences in preferences, ambitions, or forms of stimulation are independent of student performance, the impact of such factors is reduced in systems with high accountability.

Central exams are also expected to raise academic achievement in lower tracks. This is because in a tracked system with central exams, schools are held accountable for the performance of students in *all* tracks, and thus schools and teachers attempt to maximize the performance of students in all of those tracks. One outcome of central exams could therefore be that schools redistribute resources so that students in lower tracks benefit as well, and similarly schools might be motivated to hire better teachers for all tracks. These factors suggest that it could be possible for central exams to compensate for the post-placement track effects that are usually beneficial only for students in higher tracks (cf. Brunello and Checchi 2007).

In short, our proposition is that central exams may relieve the negative effect of tracking on equality of educational opportunity by mitigating the social selection and social inequality which often occur after track placement. We thus expect the association between tracking and inequality of educational opportunity to be less strong when central examinations are applied (*hypothesis 2*).

DATA AND METHODS

Data

The main source of data used for this study is the Programme of International School Assessment (PISA) for the year 2006.[7] This program, developed by the OECD, is an international study which measures the cognitive skills of 15-year-old students by means of a standardized survey. For our analysis, we linked the PISA data to country-level indicators for tracking and central examinations (see below). The sample size is N_1=244,524 (36 countries). We follow the data analysis manual of the PISA data (OECD 2009b) and use sampling weights for all our analyses. The sum of weights are equalized for each country in such a way that each country carries equal weight in the estimation. Furthermore, the relative weights of different countries are normalized to sum up to the total number of observations.

Methods

Perhaps the biggest challenge in comparative research into the relation between educational systems and student achievement concerns the unobserved heterogeneity at country-level. Countries vary in many more ways than can be assessed using quantitative indicators of educational systems. Contemporary econometric studies of international student achievement data have therefore included country fixed effects rather than a country-level random intercept (Hanushek and Wössmann 2011; Brunello and Checchi 2007). Such country fixed effects can be included if one's focus is on cross-level interaction terms (as is the case in our study, between educational institutional variables and social background), thereby meaning that the main effects of country-level variables cannot be included in the model. The cross-level interaction effects then lead to potential non-linearities in the effects of individual-level variables, which is precisely the result which we are hypothesizing.

Fixing the country-level effects have a number of advantages over “normal” multilevel regression models. First, given the non-random sample of countries, multilevel models including a country level are likely to violate the assumption of normality of the country-level residuals. Second, the sample size at country-level is small (our sample has 36 countries), and as a consequence there is little freedom at the country-level, and the estimates rely on a limited number of observations. Studies have recently shown that the low number of level-2 units mean that estimates of cross-national multilevel models are highly susceptible to deviant cases (Van der Meer, te Grotenhuis, and Pelzer 2010).

The general equation for the country fixed effect model is:

$$\begin{aligned} \mathrm{Y_i} = \beta_{0\mathrm{i}} + \beta_1 \mathrm{fem_i} + \beta_2 \mathrm{age_i} + \beta_3 \mathrm{grade_i} + \beta_4 \mathrm{immig1_i} + \beta_5 \mathrm{immig2_i} + \beta_6 \mathrm{ESCS_i} + \sum_{\mathrm{k=k\text{-}1}} \beta_{\mathrm{x}} \mathrm{D}_k \\ \beta_8 \mathrm{ESCS * track_i} + \beta_9 \mathrm{ESCS * CE_i} + \beta_{10} \mathrm{ESCS * track * CE_i} + e_{\mathrm{i}} \end{aligned} \qquad \textbf{(1)}$$

In equation 1 $\mathrm{Y_i}$ is the score on the science test for individual *i*, β_1 to β_6 are all the estimates for the individual level covariates (gender, age, grade, immigration background, and social origin, respectively), β_{x} estimates the fixed effects for countries by adding dummies (dummies D for country *k*), β_8 and β_9 are estimates for the cross-level interactions between socio-economic status and the country-level variables, β_{10}— our prime focus— is the estimate of the three-way interaction between socio-economic status, tracking and central exams, and e_{ik} is the error term. If, as our hypotheses predict, tracking generally decreases equality of educational opportunity , but does this less in systems with central examinations, β_8 should be positive and β_{10} negative.

In addition, however, the PISA data allows us to incorporate the school-level into our analysis (N_2 = 9,167 schools). By using multilevel models where we take schools into account, we can also shed light on the effects of tracking while controlling for the allocation processes of schools. Because of the large and random sample of schools in the PISA data we can add a random effect for the school-level intercept without violating any assumptions. In these multilevel models students (*i*) are nested in schools (*j*), while countries are again added as a fixed effect. The general equation for these models is:

$$\begin{aligned} \mathrm{Y_{ij}} = {} & \beta_{0_{\mathrm{ij}}} + \beta_1 \mathrm{fem_{ij}} + \beta_2 \mathrm{age_{ij}} + \beta_3 \mathrm{grade_{ij}} + \beta_4 \mathrm{immig1_{ij}} + \beta_5 \mathrm{immig2_{ij}} + \beta_6 \mathrm{ESCS_{ij}} + \\ & \sum_{\mathrm{k=k\text{-}1}} \beta_{\mathrm{x}} \mathrm{D}_k + \beta_8 \mathrm{ESCS*track_{ij}} + \beta_9 \mathrm{ESCS*CE_{ij}} + \beta_{10} \mathrm{ESCS*track*CE_{ij}} + e_{\mathrm{ij}} + u_{\mathrm{0j}} \end{aligned} \quad \textbf{(2)}$$

Equation 2 is highly similar to equation 1, except that here we allow for a random intercept at school-level *j*, and this is reflected in the school-level variance term u_{0j}. Furthermore, the model as described in equation 2 allows us to incorporate other potentially relevant school-level variables (see the 'Variables' section).

<u>Variables</u>

For the dependent variable we use student performance in science. The PISA 2006 dataset includes three cognitive variables: science, mathematics and reading. In 2006, science was the major focus of the study and was therefore “comprehensively assessed on the basis of a newly developed and elaborated framework” (OECD 2009a:

244). Technically, this means that instead of the 31 items for reading or the 48 items for mathematics the science score was based on 108 items (OECD 2009a: 2). The science score is scaled so that the OECD mean score is 500 with a standard deviation of 100 (OECD 2009a).

The PISA survey uses plausible values (Rubin 1976), which are described by Wu and Adams (2002) as a "representation of the range of abilities that students might reasonably have. […] Instead of directly estimating a student's ability, a probability distribution for a student's ability is estimated" (in OECD 2009b: 96). The OECD provides five plausible values for each domain. Even though these five items are highly correlated, averaging plausible values at the individual level leads to biased estimates (OECD 2009b: 100). We therefore estimate all models separately for each plausible value, averaged all parameters between the five models, and calculated standard errors to account for both the variance within and between plausible values (OECD 2009b).

The main independent variable at the student level is socio-economic status ("Economic, Social, and Cultural Status," ESCS), a variable created by the OECD specifically for PISA. The PISA index of ESCS includes indicators of parents' occupation, parents' education, and home resources (including both financial and cultural resources). The index maintains a mean of 0 and a standard deviation of 1 for students from OECD countries (OECD 2009a). As well as socio-economic status we also control our results for the student's gender (female=1), age and include a binary variable for immigration status (first generation, second generation and native). We also control for the grade the student is in, which is essential as PISA interviews 15-year-olds independent of grade level.

On the school-level we add four control variables: first, an indicator that distinguishes between public and private schools, second the student/teacher ratio, third, a binary variable for the location of the school (village, small town, town, city, large city), and fourth, the overall socio-economic composition of the school. The score for the latter variable is obtained by averaging students' ESCS over schools. As we argued above, school composition can be highly relevant to our research question, especially since there are indications that a higher mean socio-economic status is positively associated with student achievement and that peer effects are stronger in tracked schools (Entorf and Lauk, 2006). The descriptive statistics for all the individual and school-level variables we used in the total sample can be found in Table 1.

[Table 1 about here]

The country-level data are obtained from a macro level dataset that was composed from a variety of sources by Bol and Van de Werfhorst (2011, 2013). Our main interest here is in two country-level variables: first, the level of curricular tracking between schools, and second, the existence of central examinations. In considering the association between macro-level variables and student performance we make the assumption that educational system characteristics are exogenous, and not affected in response to student achievement.

The index of tracking which we use is a combined index of three measures that indicate the level of tracking in secondary education. First of all the age of first selection (OECD 2006: 107, Table A7.1) is used as an indicator of when the actual tracking between educational programs starts. The second indicator is the number of different curricular tracks that are available for 15-year-olds, which shows us how many tracks there are for the respondents in the PISA sample (OECD 2006: 107, Table A7.1). The final indicator we use to operationalize tracking is the how long the tracked curriculum lasts. This final indicator calculates the percentage of the total compulsory curriculum which is tracked and is calculated on the basis of data from OECD (Brunello and Checchi 2007: 799). We performed a factor analysis on these three indicators and saved the results as regression coefficients.[8] Countries with an educational system that has an average level of tracking score a zero on this index. A negative deviation signals less tracking than average, while a positive deviation signals more tracking. It is important to mention that the factor analysis refers to a larger sample of countries than is studied in the rest of this article: because a factor analysis estimates the relative position of all observations it is important to include as many countries as possible (in Bol and Van de Werfhorst 2011 2013).

The criteria for the central examinations indicator are based on the operationalization of Bishop (1997: 270). He proposes five criteria that central exams should have: (1) the diploma should have real consequences and not be merely symbolic, (2) diplomas are tested against a national standard, (3) central examinations are organized by discipline, (4) the outcome is not dichotomous (pass/fail) , and (5) they exam should be part of secondary education and should also cover most of the student population in secondary education. The data for central examinations are derived from the *European Glossary on Education*, in particular from the section on examinations, qualifications and titles (Eurydice, 2004). This report gives a summary of all examinations in a country and furthermore tells us if the exams are enforced by a national institution. If countries meet both criteria they are coded as 1. As the report only covers European countries, our data is supplemented and cross-checked with data from Wössman et al. (2009: 123). The vast majority of countries receive a value of 0 (no exams) or 1 (exams). However, following Wössman et al. (2009), four countries were given a different value (Australia, Canada, Germany and the United States), due to the fact that central examinations are held only in some states or provinces within the country. Table 2 shows the scores for all countries for both of the institutional indicators which are relevant for the present analysis.

[Table 2 about here]

RESULTS

We will first discuss the models which look at individuals as nested in countries without taking the school-level into account (following equation 1). Then we will go on to show the results from our models where we add the school-level to the analysis and control our results for relevant school characteristics (following equation 2).

Country-level fixed effects models

The null model in Table 3 shows that the country fixed effects account for almost 8% of the total variation in the students' score on the PISA science test. In Model 1, the relation between individual level control variables (female, grade year, age, immigration status, and socio-economic status) and student performance are estimated. All effects are highly significant in the direction that one would expect on the basis of previous studies. The indicators for female, age, and both first and second generation immigrants (with the status 'native' as a reference category) are negatively related to student performance.

Girls score on average 3.6 points less than boys, while first and second generation immigrants do significantly worse (23.9 and 22.9 points, respectively) than native students. Grade year is positively associated with performance as those in higher grade levels perform better. As expected, the effect of socio-economic status (ESCS index) on student achievement is positive: for every standard deviation increase in socio-economic status, the PISA score increases by 34.0 points.[9]

[Table 3 about here]

In Model 2 we add the cross-level interaction between tracking and socio-economic status. The main effect of tracking is not added to this model, since the fixed effects of countries already capture all of the between-country heterogeneity, including the heterogeneity in tracking regimes. The cross-level interaction effect between tracking and socio-economic status is significant and positive: the higher the level of tracking in an educational system, the greater the effect of social background on student performance tends to be. For each 1 point increase on the scale of tracking, the predicted effect of socio-economic status on performance in the science test increases by 1.4 points. These findings are in line with earlier empirical studies of tracking (e.g., Brunello and Checchi 2007; Van de Werfhorst and Mijs 2010), and allow us to confirm *hypothesis 1a*: when educational systems are more tracked, social class background correlates more strongly with student achievement.

Before we discuss the three-way interaction, let us first model the cross-level interaction of central examinations with socio-economic status in Model 3. The cross-level interaction between central examinations and socio-economic status is significantly negative, indicating that in countries with central examinations there is greater equality in student performance across different socio-economic backgrounds. One likely explanation for this finding as suggested earlier, is, that when schools are made more accountable for their results, this increases their incentives to select students on the basis of their achievement rather than who their parents are. However, we do not attach too much significance to this finding, as in the later models (see Table 5) we show that the negative interaction between social class and central examinations is not robust once school-level control variables are included in the calculations.

In Model 4 we add the variable that is the central focus of the article: the three-way interaction between socio-economic background, central exams, and tracking. By focusing on this three-way interaction we are able to test the hypothesis that although socio-economic status has a stronger effect on student performance in strongly tracked educational systems, this effect might be mitigated if the system also has central exams. For this model, we find a negative effect of -2.0 which is highly significant. The negative effect of the three-way interaction thus shows that the effect of tracking on inequality in student performance by socio-economic background does indeed decrease when central exams are applied.

Indeed, if we compare the interaction between tracking and socio-economic background in the different models, we see that this effect for model 4 is almost double that of Models 2 and 3. For each 1 point increase on the tracking index, the effect of ESCS on student performance increases by 2.7 points. This estimate refers to the magnifying effect of tracking on inequality in systems *without* central examinations (i.e. when central examinations = 0). Thus if there are no central examinations at the end of secondary education, the magnifying effect of tracking on the relation between socio-economic background and student performance does indeed become even larger than previously thought. However, when central exams are implemented this negative effect of tracking is offset to a great extent. This is shown best in an example.

Following the lead of Aiken and West (1991), we can estimate the marginal effect *P* of ESCS on performance for varying scores of tracking and central exams by using the following equation:

$$P = 34.8+(2.7*\text{tracking})+(-1.2*\text{CE})+(-2.0*\text{tracking}*\text{CE}) \quad \textbf{(3)}$$

This formula shows that the positive effect of tracking is largely offset by the implementation of central exams. In a highly tracked country (measuring 1.5 on the tracking index) the predicted marginal effect of ESCS on school performance would be 34.7 instead of 38.9, a difference of more than 4 points. Moreover, the three-way interaction shows that this difference is significant. In countries with central examinations the relationship between tracking and inequality in learning by socio-economic position is largely attenuated. This finding therefore confirms *hypothesis 2*.

Figure 1 depicts the above calculation, and the main finding of Table 3. The graph shows that, in the absence of central examinations, educational institutions with tracking are associated with larger social inequalities in learning. However, if schools are subject to central examinations, even if there is tracking, they are still subject to a higher level of accountability and students' achievement is therefore less determined by their socio-economic position.

[Figure 1 about here]

<u>Adding the school-level to the analysis</u>

As was discussed earlier section in this paper, it is important to take the school-level into account, however this also changes the interpretation of the country-level effect of tracking (see *hypothesis 1b*). Schools play an intermediary role here, they are located between educational systems on the one hand, and students' opportunities for educational performance on the other. In many countries, the “tracking effect” is largely caused by the allocation of students to schools of different educational “levels”. However, it is important to see if our main result—the negative three-way interaction between tracking, central exams and socio-economic background—is also confirmed in a model which allows us to fully exploit the nested structure of the PISA data.

Table 4 presents models that are comparable to Table 3, with the exception that the school-level is now added to our analysis. In the null model of Table 4 we can see that schools make up 32.1 percent of the total variance in student performance on the PISA science test (ICC). Although adding the school-level to our analysis does not change the direction or significance of the individual level effects, the effect sizes are different. In Model 1 we see that girls are predicted to do even worse (-9.2) than was estimated in the same model in Table 4. Moreover, the effect of ESCS on student performance decreases to roughly two-thirds of the original estimate when we take the school-level into account.

[Table 4 about here]

When we consider the interaction effects between socio-economic background and our contextual variables, even more important implications arise. Now, the negative interaction between socio-economic status and central examinations remains, although the effect size is larger. However, the most important difference between Table 3 and Table 4 is that the interaction effect between socio-economic background and tracking changes direction. In the models where we nest students in schools this interaction is now significantly negative: for each one unit increase on the scale of tracking, the effect of ESCS on student performance on the PISA science test decreases by 6.4 points. We therefore confirm our *hypothesis 1b*.

If we take the average student performance of schools into account by adding a random intercept across schools, the association between socio-economic background and student performance changes. This finding can be explained by the fact that in educational systems with explicit tracking (most obviously in systems such as the German one which have different curricular tracks in separate schools, but also in countries with less tracked systems such as Belgium) pupils are allocated to different schools and/ or locations based on their earlier performance (and thus also on parental background). It is important to note that while the interaction effect of tracking and socio-economic background is negative in our three-level models, we are nonetheless not arguing that tracking can decreases inequality in student achievement along socio-economic lines. Rather, our results merely show that the effect that tracking has on the relation between social class and student performance is mainly due to the placement of students in different schools (cf. Dronkers, Van der Velden, and Dunne 2012).

The main finding of Table 3 is the three-way interaction between socio-economic background, tracking, and central examinations, and this is also confirmed in Table 4. Even when the role of schools is taken into consideration, in systems without central examinations tracking is still more detrimental to students' equal opportunities across social backgrounds than it is in systems with central examinations. Figure 2 depicts the marginal effect of ESCS for changing values of tracking and central exams once again, this time based on Table 4. The figure shows that the association between tracking and the marginal effect of ESCS on student performance is smaller in educational systems which have central exams.

[Figure 2 about here]

Adding the school-level also allows to control our findings for school-level characteristics (Table 5). State schools (see Model 5) are very strongly associated with poorer student performance, meaning that in general students tend to perform better in private schools (cf. Dronkers 2008). The student/teacher ratio is positively related to student performance, suggesting that students in schools with bigger classes perform better. Although this might be a surprising finding, it is actually in line with earlier studies that tended to find counterintuitive signs of the coefficient of class size (Hanushek and Wössmann 2011). Moreover, in Model 8 we find that this effect is not robust for the inclusion of other school-level controls (see also Fuchs and Wössmann 2007).

In Model 7 we add dummy variables for the location of the school, and of these 3 out of 4 show significant effects. In this model we can also see that by controlling for several school-level characteristics, the negative interaction between central exams and ESCS loses its significance. Both the two-way interaction between tracking and ESCS and the three-way interaction give relatively stable and significant estimates, and these effects persist even when we add a final control variable for the socio-economic composition of the school. Thus in Model 8 we see that, perhaps unsurprisingly, the social composition of schools has a large effect on student performance (64.5).

[Table 5 about here]

Most important is that our main conclusion is not affected when we control for several relevant school characteristics. The effect of socio-economic background on student performance is still lower in tracked educational systems, and the three-way interaction persists, indicating that central exams are indeed associated with a smaller effect of tracking on the relation between socio-economic background and educational achievement.

CONCLUSION

By now many studies have shown that we find greater social inequality in the achievement levels of students in educational systems where they are separated early into different curricular tracks than in comprehensive systems (Brunello and Checchi 2007; Van de Werfhorst and Mijs 2010; Marks 2005). However, what such studies have often ignored is the question of whether the relationship between tracking and the inequality in learning by socio-economic background is affected by other institutional characteristics of educational systems. The question of whether other institutional features potentially counterbalance this relation had not previously been explored adequately. In this paper we sought to fill this gap, and studied one aspect of educational systems which has a profound influence on how tracking is related to inequality, namely central examinations.

Using data from PISA 2006, we confirmed earlier findings that inequality in learning due to students' socio-economic background is larger in educations which are strongly tracked as compared to comprehensive systems. However, when we further differentiate the relationship between tracking and socio-economic inequality, our study shows that the negative impact of tracking pupils in different curricula is also affected by whether or not countries have implemented central examinations within secondary educational institutions. In societies where central examinations are prevalent in the secondary school system, the relationship between the level of tracking and the level of inequality by socio-economic status is attenuated.

The inclusion of the school-level in our analysis demonstrated that the measured and unmeasured characteristics of schools also affect what impact tracking has on the importance of social background for learning. Differences between schools that are a result of between-school tracking—selection and track effects—are thus also accounted for. Consequently the indirect effect of social background via measured and unmeasured school characteristics on achievement is no longer visible. When the school-level is taken into account, the total effect of parental background on achievement is underestimated, something we substantiated by the large effect of socio-economic composition of schools on student performance.

Even in the analyses which incorporated the school-level, a negative three-way interaction between socio-economic background, tracking and central exams was found. We explained this finding by suggesting that schools and students are more easily held accountable for their performance if they have to take part in central examinations (cf. Wössmann et al. 2009). Such exams offer incentives to perform for both parties, and this is likely to particularly affect schools which provide the lower (pre-)vocational tracks which have a higher proportion of children from lower socio-economic backgrounds . Children from disadvantaged backgrounds therefore benefit more from the educational system than children from higher socio-economic backgrounds, who benefit more from their family resources. In countries with centralized examinations, therefore, the lower tracks are less evidently considered as “waste bins for the untalented” than they are in countries where schools are less easily held accountable for performance.

With this result we are able to build on the findings of Ayalon and Gamoran (2000), expanding from their focus on Israel and the United States to a much wider set of countries. As they suggested, without formal testing, the detrimental effects of tracking on equality are much more pronounced if there are no national regulations that govern the accountability of schools.

These findings are highly relevant for educational policy in the sense that they shed new light on the relation between tracking and educational inequality. The OECD (2007: 14) have suggested that “limiting early tracking and streaming and postpone academic selection” is an important step to promote equity in education. However our work shows that the negative impact of tracking can be mitigated by incorporating “standardizing” institutions to counteract the strong impact of parents on the placement of students in different tracks at the beginning of their school career. Alongside the move to reform educational systems towards more comprehensive education, another way to combat inequality of educational opportunity is to incorporate central examinations at various moments before and during the secondary school years. This policy implication of our finding is exceptionally important, given that the OECD (2007) is critical of countries with strongly tracked educational systems. Their criticism now requires some adjustment, as our study shows that it should particularly be addressed towards countries that have tracking but do not have any central examinations.

Endnotes

[1] 'Secondary education' means the education which follows the basic programs of the primary level and includes the final stage of compulsory education. Often subjects are taught by more specialized teachers who mainly conduct classes in their field of specialization. Typically, cross-national research has examined tracking mostly in secondary education (e.g. Marks 2005; Schütz et al.2008), as tracking in higher education is less institutionalized (but see Shavit et al., 2007). Given that we are studying secondary school achievement, we also focus on curricular tracking between schools within secondary education.

[2] Tracked educational systems can differ in their levels of 'track mobility', the possibility for a student to move from one track to another (Kerckhoff 2001). While we acknowledge that track mobility might potentially offset some of the negative effects of tracking, data limitations do not allow us to incorporate this dimension into our analysis.

[3] Alongside social background, geographical location can also influence a student's access to school and hence the allocation of students. Unfortunately our data does not allow us to see how far between-school homogeneity or heterogeneity is a result of residential segregation.

[4] A presupposition here is that where central exit examinations are applied, the test results of each school are made public. While this is true for most cases, it should be noted that differences in the level of transparency of school performance may affect the impact of central exit exams on student performance.

[5] A possible relation between the availability of information about performance across schools and an increased incentive to perform due to competition is subject to the

possibility of parents to choose the school of their preference. At the same time, the educational system and the geographical location may limit the schools on offer.

[6] This suggested positive effect of central examinations on educational performance has not gone without criticisms. A review study conducted under auspices of the National Research Council in fact concluded that test-based incentive programmes have failed to increase student performance in the United States (Hout and Elliott 2011). Opponents of central examinations claim that exams are not an appropriate way to measure performance since it is still teachers who will prepare their students for the tests and testing therefore leads to unethical test preparation (Jones et al., 2003). Moreover, it is argued that "teaching to the test" has negative side-effects, for example leading teachers and students to pay less attention to content that is not tested and have less engagement with critical thinking (Amrein and Berliner, 2003).

[7] Replication of our study with the more recent PISA 2009 data yields results which are highly similar to those presented in this article.

[8] One factor was extracted from the data, which has an eigenvalue of 1.76. More information on the procedure can be found in Bol and Van de Werfhorst (2011, 2013).

[9] It is known that there is a link between immigrant status and ESCS. In order to rule out the possibility that our findings are biased by this interaction, we ran additional multilevel random intercept regression models which included the interaction between immigrant status and ESCS. In these models, we found that a positive interaction between immigrant status and ESCS. This shows that the negative main effect of immigrant status on science performance decreases for immigrant students with a higher socio-economic status. These results are available upon request.

References

Aiken, L.S., and West, S. (1991). *Multiple regression: Testing and interpreting interactions.* Thousand Oaks: Sage.

Allmendinger, J. (1989). Educational systems and labor market outcomes. *European Sociological Review*, *5*(3), 231-250.

Amrein, A. L., and Berliner, D. C. (2003). The effects of high-stakes testing on student motivation and learning. A research report. *Educational Leadership, 60*(5), 32-38.

Ayalon, H., and Gamoran, A. (2000). Stratification in academic secondary programs and educational inequality in Israel and the United States. *Comparative and International Education Society*, *44*(1), 54-80.

Bishop, J. (1997). The effect of national standards and curriculum-based exams on achievement. *The American Economic Review*, *87*(2), 260-264.

Bol, T., & Van de Werfhorst, H.G. (2011). Measuring Educational Diversity: Tracking, Vocational Orientation, and Standardization. *AMCIS Working Paper, December 2011*.

Bol, T., & Van de Werfhorst, H.G: (2013). Educational Systems and the Trade-off Between Labor Market Allocation and Equality of Educational Opportunity. *Comparative Education Review* 57(2).

Boudon, R. (1974). *Education, opportunity and social inequality*. New York: Wiley.

Breen, R., and Goldthorpe, J. H. (1997). Explaining educational differentials. Towards a formal rational action theory. *Rationality and Society*, *9*(3), 275-305.

Breen, R., and Jonsson, J. (2005). Inequality of opportunity in comparative perspective: Recent research on educational attainment and social mobility. *Annual Review of Sociology, 31*, 223-243.

Breen, R., and Yaish, M. (2006). Testing the Breen-Goldthorpe model of educational decision making. In S.L. Morgan, D.B. Grusky, and G.S. Fields (Eds.) *Mobility and inequality: Frontiers of research in sociology and economics* (pp. 232-258). Palo Alto, CA: Stanford University Press.

Brunello, G., and Checchi, D. (2007). Does school tracking affect inequality of opportunity? New international evidence. *Economic Policy*, *22*(52), 781-861.

Chiang, H. (2009). How accountability pressure on failing schools affects student achievement. *Journal of Public Economics*, *93*(9–10), 1045–1057.

Coleman, J. S. (1966). *Equality of Educational Opportunity*. National Center for Educational Statistics (DHEW/OE), Washington, DC.

Cross, C. (2004). *Putting the pieces together; Lessons from comprehensive school reform research*. The National Clearinghouse for Comprehensive School Reform (NCCSR).

Dronkers, J., 2008. Education as the backbone of inequality - European education policy: constrains and possibilities. pp. 50-135 in *Social Democracy and Education. The European Experience*, edited by F. Becker, K. Duffek & T. Mörschel. Berlin /Wien/ Amsterdam: Friederich Ebert Stiftung / Karl Renner Institut / Wiardi Beckman Stichting

Dronkers, J., Van der Velden, R., & Dunne, A. (2012). Why are Migrant Students Better Off in Certain Types of Educational Systems or Schools than in Others? *European Educational Research Journal* 11(1): 11-44.

Dunne, Allison. 2010. Dividing Lines: Examining the Relative Importance of Between and Within- School Differentiation during lower Secondary Education. *PhD thesis, European University Institute.*

Duru-Bellat, M., and Kieffer, A. (2000). Inequalities in educational opportunities in France: Educational expansion, democratization or shifting barriers? *Journal of Education Policy*, *15*(3), 333-352.

Entorf, H., and Lauk, M. (2006). *Peer effects, social multipliers and migrants at school: An international comparison*. IZA Discussion Paper 2182.

Erikson, R., and Jonsson, J. (1996). *Can education be equalized? The Swedish case in comparative perspective.* Boulder: Westview press.

Eurydice. (2004). *European Glossary on Education. Volume 1 – Second Edition. Examinations, Qualifications and Titles*. Brussels: Eurydice.

Figlio, D., and Page, M. (2002). School choice and the distributional effects of ability tracking: does separation increase inequality? *Journal of Urban Economics*, *51*(3), 497-514.

Fuchs, T., and Wössmann, L. (2007). What accounts for international differences in student performance? A re-examination of using PISA data. *Empirical Economics*, *32*, 433–464.

Gamoran, A. (1996). Curriculum standardization and equality of opportunity in Scottish secondary education: 1984-90. *Sociology of Education*, *69*(1), 1-21.

Gamoran, A., and Weinstein, M. (1998). Differentiation and opportunity in restructured schools. *American Journal of Education*, *106*(3), 385-415.

Hallinan, M. (1994). Tracking: From theory to practice. *Sociology of Education*, *67*(2), 79–84.

Hanushek, E. A., and Wössmann, L. (2011). The Economics of International Differences in Educational Achievement. In E. A. Hanushek, S. Machin, and L. Wössmann (Eds.), *Handbooks in Economics* (Vol. 3, pp. 89–200). The Netherlands: North-Holland/Elsevier.

Hanushek, E., and Wössmann, L. (2006). Does educational tracking affect performance and inequality? Differences- in-differences evidence across countries. *The Economic Journal, 116*(3), C63-C76.

Horn, D. (2009). Age of selection counts: a cross-country analysis of educational institutions*. Educational Research and Evaluation, 15*(4), 343-366.

Hout, M., and Elliott, S. W. (Eds.). (2011). *Incentives and Test-Based Accountability in Education. Report of the Committee on Incentives and Test-Based Accountability in Public Education, National Research Council*. Washington D.C.: National Academies Press.

Jones, M. G., Jones, B. D., and Hargrove, T. Y. (2003).The unintended consequences of high stakes testing. Lanham, MD: Rowman and Littlefield Publishers Inc.

Kerckhoff, A. (2001). Education and social stratification processes in comparative perspective. *Sociology of Education*, 74(extra issue), 3-18.

LeTendre, G. K., Hofer, B. K., and Shimizu, H. (2003). What is tracking? cultural expectations in the United States, Germany, and Japan. *American Educational Research Journal, 40*(1), 43-89.

Marks, G. (2005). Cross-national differences and accounting for social class inequalities in education. *International Sociology*, *20*(4), 483–505.

Marks, G. (2006). Are between- and within-school differences in student performance largely due to socio-economic background? Evidence from 30 countries. *Educational Research*, *48*(1), 21-40.

Meghir, C., and Palme, M. (2005). Educational Reform, Ability, and Family Background. *American Economic Review, 95*(1), 414-424.

Minter Hoxby, C. (2000). *Peer effects in the classroom: learning from gender and race variation*. NBER Working Paper 7867. Cambridge, MA: National Bureau of Economic Research.

Oakes, J. (2005). *Keeping track, how schools structure inequality*. New Haven: Yale University Press.

OECD. (2006). *Education at a glance. OECD indicators 2006*. Paris: Organisation for Economic Co-operation and Development.

OECD. (2007). *No More Failures: Ten Steps to Equity in Education*. Paris: Organisation for Economic Co-operation and Development.

OECD. (2009a). *Pisa 2006 Technical report*. Paris: Organisation for Economic Co-operation and Development.

OECD. (2009b). *Pisa 2006 Data Analysis Manual (SPSS)*. Paris: Organisation for Economic Co-operation and Development.

Pekkarinen, T., Uusitalo, R., and Kerr, S. (2009). School tracking and development of cognitive skills. *IZA Discussion Paper No. 4058*.

Richer, S. (1976). Reference-group theory and ability grouping: A convergence of sociological theory and educational research. *Sociology of Education, 49*(1), 65-71.

Rubin, D. B. (1976). Inference and missing data. *Biometrika*, *63*(3), 581–592.

Schütz, G., Ursprung, H., and Wössmann, L. (2008). Education policy and equality of opportunity. *Kyklos*, *61*(2), 279-308.

Shavit, Y., and Blossfeld, H. (1993). *Persistent inequality; changing educational attainment in 13 countries*. Boulder, CO: Westview press.

Shavit, Yossi, and Walter Müller. 1998. *From school to work: A comparative study of educational qualifications and occupational destinations*. Oxford: Oxford University Press.

Shavit, Y., Arum, R., and Gamoran, A. (2007). *Stratification in Higher Education: A Comparative Study*. Stanford, CA: Stanford University Press.

Van de Werfhorst, H.G, & Mijs, J.J.B. (2010). Achievement inequality and the institutional structure of educational systems: A comparative perspective. *Annual Review of Sociology*, 36, 407-428.

Van der Meer, T., Te Grotehuis, M., and Pelzer, B. (2010). Influential Cases in Multilevel Modeling. *American Sociological Review, 75*(1), 173-178.

Van Houtte, M., and Stevens, P. A. J. (2009). Study Involvement of Academic and Vocational Students: Does Between-School Tracking Sharpen the Difference? *American Educational Research Journal*, *46*(4), 943–973.

Wilkinson, I., Parr, J., Fung, I., Hattie, J., and Townsend, M. (2002). Discussion: Modelling and maximizing peer effects in school. *International Journal of Educational Research, 37*(4), 521-535.

Wössmann, L. (2003). Schooling resources, educational institutions and student performance: The international evidence. *Oxford Bulletin of Economics and Statistics*, *65*(2), 117–171.

Wössmann, L., Luedemann, E., Schütz, G., and West, M. (2009). *School accountability, autonomy, and choice around the world*. Cheltenham, UK: Edward Elgar.

Wu, M., and Adams, R. (2002, April). *Plausible values: why are they important?* Paper presented at the International Objective Measurement Workshop, New Orleans, LA.

Tables

Table 1. Descriptive statistics for all used variables.

	Mean	Standard Deviation	Minimum score	Maximum score
Individual level variables				
Plausible values				
Plausible value science 1	506.785	97.879	71.838	920.482
Plausible value science 2	506.963	97.936	23.443	924.212
Plausible value science 3	506.761	97.876	70.066	897.917
Plausible value science 4	506.757	98.002	39.108	883.183
Plausible value science 5	506.852	97.914	22.790	952.187
Female	0.500	0.500	0.000	1.000
Grade	9.694	0.711	7.000	13.000
Age	15.783	0.291	15.170	16.330
Immigrant status				
Native citizen	0.901	0.298	0.000	1.000
2^{nd} generation immigrant	0.050	0.219	0.000	1.000
1^{st} generation immigrant	0.048	0.215	0.000	1.000
ESCS-index	0.041	0.939	-5.670	3.349
School-level variables				
Public school	0.828	0.377	0.000	1.000
Student/teacher ratio	12.705	4.627	0.727	48.333
School location				
Village	0.102	0.303	0.000	1.000
Small town	0.250	0.433	0.000	1.000
Town	0.343	0.475	0.000	1.000
City	0.227	0.419	0.000	1.000
Large city	0.078	0.268	0.000	1.000
Mean school ESCS	0.047	0.548	-2.953	1.754

NOTE. The descriptive statistics for the individual level variables are calculated using our empirical sample of PISA 2006 (N=244,534). The descriptive statistics for the school-level variables are calculated by using only those observations which had no missings for either of the school variables (N=200,641).

Table 2. Descriptive statistics for country-level variables.

	Tracking index	Central exams
Australia	-1.078	0.81
Austria	1.751	0
Belgium	1.041	0
Bulgaria	-0.075	1
Canada	-1.315	0.51
Czech Rep.	1.671	1
Denmark	-0.930	1
Finland	-0.930	1
France	-0.477	1
Germany	1.789	0.44
Great Britain	-1.078	1
Greece	-0.477	0
Hong Kong	-0.204	1
Hungary	1.297	1
Iceland	-0.875	1
Ireland	-0.128	1
Israel	-0.127	1
Italy	0.184	1
Japan	-0.477	1
Korea Rep.	0.104	1
Latvia	-0.477	1
Liechtenstein	0.552	1
Luxembourg	0.755	1
Netherlands	0.971	1
New Zealand	-0.546	1
Norway	-1.078	1
Poland	-0.043	1
Portugal	-0.043	1
Russia	-0.250	1
Slovakia	1.059	0
Slovenia	0.764	1
Spain	-0.803	0
Sweden	-1.058	0
Switzerland	-0.024	0
Turkey	1.110	1
United States	-1.315	0.09

Table 3. Regression models with country fixed effects. Dependent variable: performance on PISA science test.

	Model 0	Model 1	Model 2	Model 3	Model 4
Individual level					
Country fixed effects	*yes*	*yes*	*yes*	*yes*	*yes*
Female		-3.605**	-3.603**	-3.600**	-3.595**
		(0.482)	(0.482)	(0.482)	(0.482)
Grade		41.516**	41.451**	41.430**	41.445**
		(0.521)	(0.521)	(0.521)	(0.521)
Age		-6.040**	-6.015**	-6.003**	-6.003**
		(0.884)	(0.884)	(0.885)	(0.884)
Native		ref.	ref.	ref.	ref.
Second generation immigrant		-23.860**	-23.702**	-23.667**	-23.669**
		(1.221)	(1.220)	(1.221)	(1.221)
First generation immigrant		-22.861**	-22.801**	-22.749**	-22.755**
		(1.516)	(1.515)	(1.517)	(1.517)
Socio-economic status (ESCS index)		33.993**	33.961**	34.890**	34.837**
		(0.276)	(0.274)	(0.458)	(0.459)
Interactions					
ESCS*Tracking			1.449**	1.451**	2.657**
			(0.281)	(0.277)	(0.441)
ESCS*Central exams				-1.246*	-1.166*
				(0.581)	(0.579)
ESCS*Central exams*Tracking					-2.043**
					(0.650)
Constant	529.259**	204.719**	205.304**	205.327**	205.066**
	(0.958)	(13.294)	(13.293)	(13.293)	(13.293)
R^2	0.079	0.258	0.258	0.258	0.259
Observations	244,534	244,534	244,534	244,534	244,534

SOURCE. PISA 2006, own calculations with the data for 36 countries (see table 2).

NOTE. Standard error in brackets. All standard errors calculated by taking into account both the variance between and within plausible values. Sampling weights were used in all analyses.

**p<0.01, *p<0,05, two-tailed tests.

Table 4. Multilevel linear regression models with a random school intercept and fixed country effects. Dependent variable: performance on PISA science test.

	Model 0	Model 1	Model 2	Model 3	Model 4
Individual level					
Country fixed effects	*yes*	*yes*	*yes*	*yes*	*yes*
Female		-9.177**	-9.350**	-9.344**	-9.337**
		(0.510)	(0.508)	(0.508)	(0.508)
Grade		36.735**	36.720**	36.697**	36.713**
		(0.607)	(0.614)	(0.616)	(0.617)
Age		-4.348**	-4.304**	-4.289**	-4.302**
		(0.766)	(0.767)	(0.767)	(0.767)
Native		ref.	ref.	ref.	ref.
Second generation immigrant		-19.747**	-20.167**	-20.120**	-20.132**
		(1.249)	(1.283)	(1.287)	(1.291)
First generation immigrant		-20.703**	-20.780**	-20.709**	-20.750**
		(1.505)	(1.516)	(1.518)	(1.521)
Socio-economic status (ESCS index)		19.890**	19.780**	21.627**	21.626**
		(0.344)	(0.323)	(0.509)	(0.512)
Interactions					
ESCS*Tracking			-6.385**	-6.381**	-4.567**
			(0.322)	(0.321)	(0.521)
ESCS*Central exams				-2.480**	-2.480**
				(0.673)	(0.674)
ESCS*Central exams*Tracking					-3.039**
					(0.759)
Constant	526.203**	229.083**	227.343**	227.373**	227.282**
	(2.463)	(12.129)	(12.108)	(12.103)	(12.100)
Σu (school)	2852.274**	1916.974**	1969.174**	1968.178**	1967.989**
	(27.175)	(20.053)	(20.896)	(20.912)	(20.965)
Σe	6011.630**	5465.421**	5441.425**	5440.843**	5440.030**
	(22.566)	(20.556)	(20.526)	(20.547)	(20.553)
ICC (school)	0.321	0.260	0.266	0.266	0.266
-2LL	2,841,881	2,816,349	2,815,500	2,815,471	2,815,434
Observations	244,534	244,534	244,534	244,534	244,534
Number of schools	9,167	9,167	9,167	9,167	9,167

SOURCE. PISA 2006, own calculations with the data for 36 countries (see table 2).
NOTE. Standard error in brackets. All standard errors calculated by taking into account both the variance between and within plausible values. Sampling weights were used in all analyses.
**p<0.01, *p<0,05, two-tailed tests.

Table 5. Multilevel linear regression models with school-level controls. Dependent variable: performance on PISA science test.

	Model 5	Model 6	Model 7	Model 8
Individual level				
Country fixed effects	*yes*	*yes*	*yes*	*yes*
Female	-9.331**	-9.253**	-8.907**	-9.018**
	(0.544)	(0.554)	(0.563)	(0.561)
Grade	36.617**	36.530**	37.495**	36.349**
	(0.649)	(0.661)	(0.700)	(0.705)
Age	-3.725**	-3.530**	-4.071**	-3.716**
	(0.803)	(0.824)	(0.844)	(0.845)
Native	ref.	ref.	ref.	ref.
Second generation immigrant	-20.303**	-20.675**	-25.280**	-25.028**
	(1.416)	(1.472)	(1.615)	(1.625)
First generation immigrant	-21.120**	-20.781**	-26.789**	-26.522**
	(1.636)	(1.706)	(1.797)	(1.794)
Socio-economic status (ESCS index)	21.364**	21.078**	20.626**	18.378**
	(0.518)	(0.529)	(0.528)	(0.539)
Interactions				
ESCS*Tracking	-4.466**	-4.309**	-4.210**	-3.732**
	(0.526)	(0.544)	(0.543)	(0.547)
ESCS*Central exams	-2.096**	-1.849**	-0.428	0.533
	(0.695)	(0.709)	(0.710)	(0.711)
ESCS*Central exams*Tracking	-3.031**	-3.211**	-3.869**	-4.321**
	(0.775)	(0.799)	(0.792)	(0.793)
School-level				
Public school	-10.663**	-9.308**	-9.013**	11.452**
	(1.826)	(1.893)	(1.897)	(1.655)
Student/teacher ratio		1.370**	1.153**	0.187
		(0.196)	(0.197)	(0.146)
School located in village			ref.	ref.
School located in small town			2.032	-9.560**
			(1.709)	(1.675)
School located in town			4.875**	-14.763**
			(1.775)	(1.698)
School located in city			7.224**	-20.363**
			(1.931)	(1.823)
School located in big city			11.067**	-19.721**
			(2.874)	(2.453)
Mean school ESCS				64.487**
				(1.354)
Constant	195.917**	172.457**	170.308**	181.894**
	(13.105)	(13.834)	(14.153)	(13.837)
Σu (school)	1940.365**	1953.321**	1922.736**	1193.054**
	(22.061)	(22.750)	(22.929)	(14.478)
Σe	5411.383**	5378.316**	5390.649**	5390.674**
	(22.155)	(22.560)	(23.205)	(23.232)
ICC (school)	0.264	0.266	0.263	0.181
-2LL	2,536,608	2,418,417	2,336,634	2,333,879
Observations	218,161	205,491	200,461	200,461
Number of schools	8,254	7,743	7,576	7,576

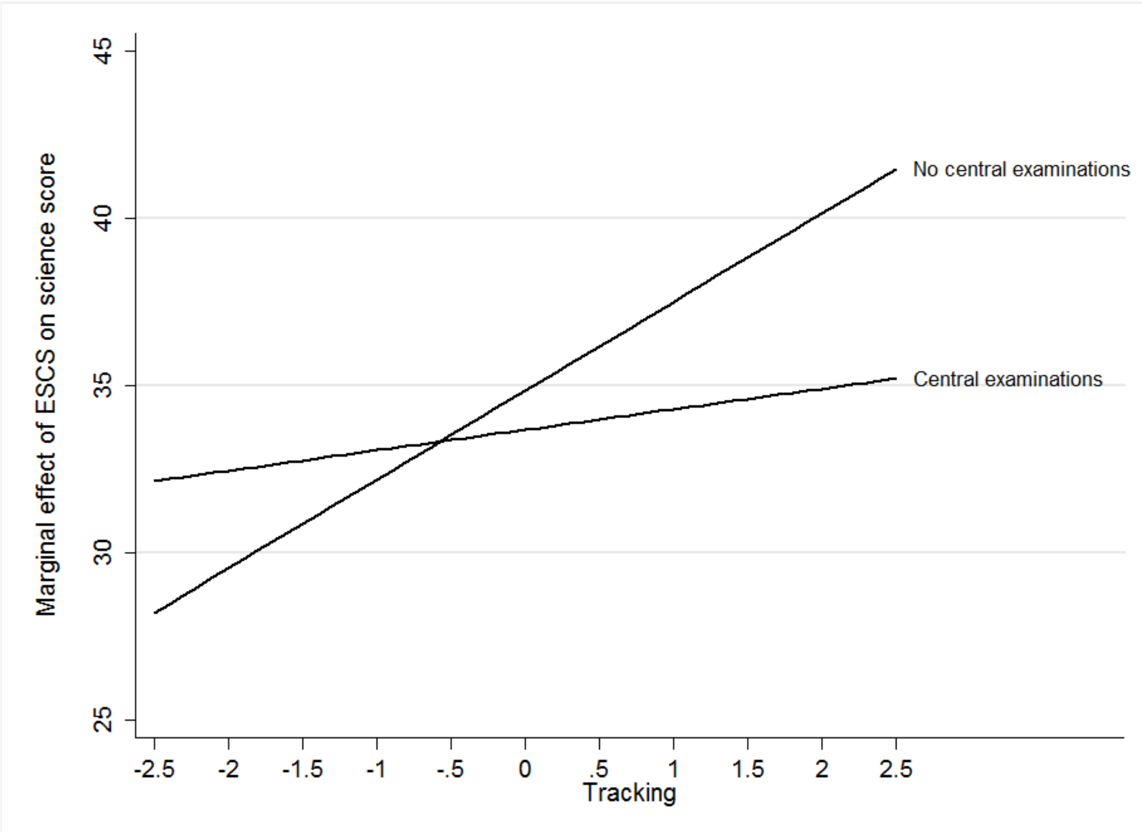
SOURCE. PISA 2006, own calculations with the data for 32 countries (see Table 2).

NOTE. Standard error in brackets. All standard errors calculated by taking into account both the variance between and within plausible values. Sampling weights were used in all analyses. In models 5, 6,7 and 8 the number of observations is not constant due to missing school-data for Australia, Bulgaria, France, Hong Kong and several schools. The presented results are the same when we use a constant sample of 200,461.

**p<0.01, *p<0,05, two-tailed tests.

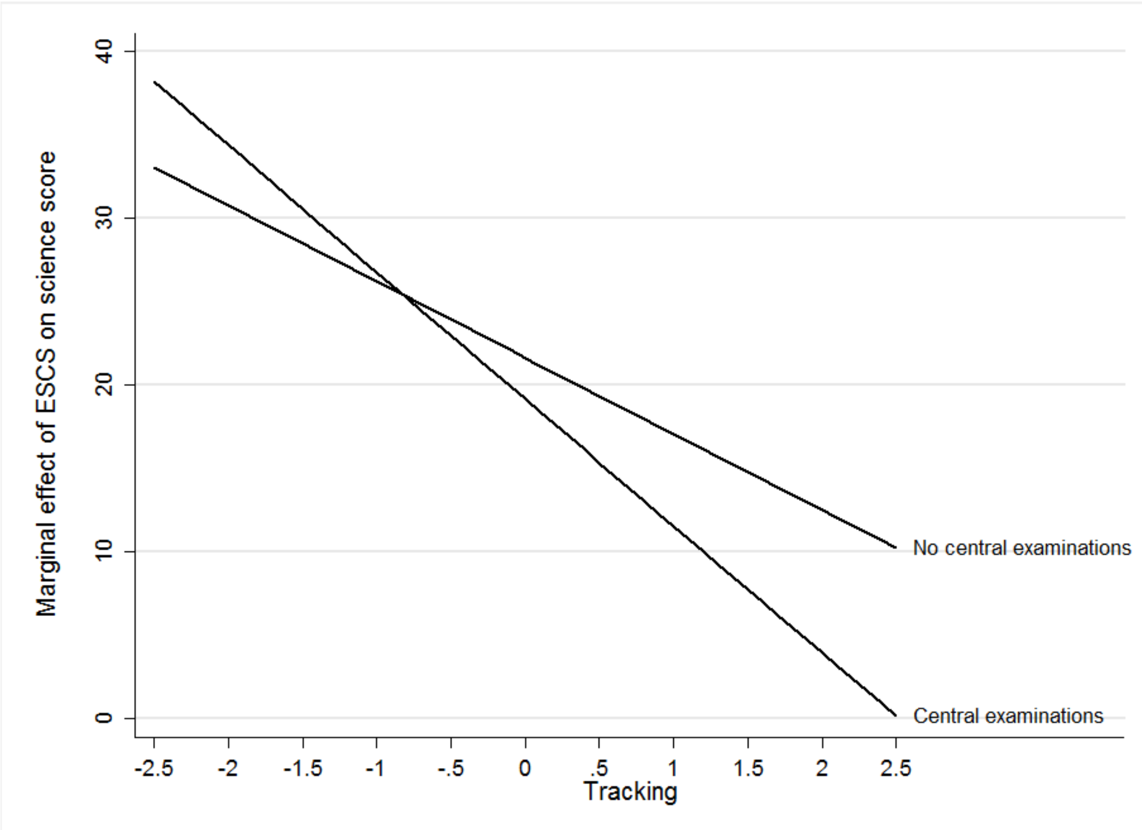
Figures

Figure 1. How tracking and central exams work together: summarizing the two-level models



NOTE. Based on Model 4, Table 3. There is a significant difference between the slopes for tracking in the absence of central examinations and presence of central examinations. ESCS refers to the PISA measure of socio-economic status.

Figure 2. How tracking and central exams work together: summarizing the three-level models



NOTE. Based on Model 4, Table 4. There is a significant difference between the slopes for tracking in the absence of central examinations and the presence of central examinations. ESCS refers to the PISA measure of socio-economic status.