Curricular Tracking and Central Examinations

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

Thijs Bol, *University of Amsterdam*
Jacqueline Witschge, *University of Amsterdam*
Herman G. Van de Werfhorst, *University of Amsterdam*
Jaap Dronkers, *University of Maastricht*

Tracked educational systems are associated with 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 2009 PISA survey, we study how central examinations affect the association between tracking and inequality. We find that parental socioeconomic status 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, which (1) encourages schools to allocate students to tracks on the basis of more objective indicators and (2) makes it likely for schools to invest more in lower-track students. Thus, central exams attenuate the stronger impact of parental status on children's performance in tracked educational systems.

Introduction

A significant body of comparative research has shown that the level of inequality in educational outcomes differs across Western societies. In addition to 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 coun-

Direct correspondence to: Thijs Bol (t.bol@uva.nl) at the University of Amsterdam, Oudezijds Achterburgwal 185, 1012 DK, Amsterdam, The Netherlands. *This research was supported by two grants from the Netherlands Organization for Scientific Research (NWO grants 542-07-002 and 411-10-920). Earlier versions of this paper have been presented at the ISA-RC28 conference in Essex (2011), the SIM-life conference in Mannheim (July 2011), the SUNSTRAT workshop at the Swedish Institute of Social Research (April 2012), and the Onderwijs Research Dagen in Wageningen (2012). The authors thank participants at those meetings, and anonymous reviewers of this journal, for valuable comments.*

Social Forces 92(4) 1545–1572, June 2014
doi: 10.1093/sf/sou003
Advance Access publication on 1 March 2014

tries (Allmendinger 1989; Shavit and Müller 1998; Kerckhoff 2001). The aspect of educational systems that is argued to have the strongest effect on students' achievement is curricular tracking, 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 that allocate students to a large number of different educational tracks at an early age are found to increase social inequalities, primarily because tracking magnifies the impact of socio-economic status on educational achievement and attainment (Marks 2005; Horn 2009). These results are confirmed by various studies that show that inequality has been reduced in countries that have moved from a tracked to a comprehensive educational system (Gamoran 1996; Gamoran and Weinstein 1998; Duru-Bellat and Kieffer 2000; Meghir and Palme 2005; Pekkarinen, Uusitalo, and Kerr 2009).

However, studies addressing the effect of curricular tracking on educational inequality have generally overlooked additional institutional characteristics that influence the effect of social origin on educational achievements. Such studies have 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). In the current study, we aim to fill this research gap. We argue that the extent to which educational systems are nationally standardized (Erikson and Jonsson 1996; Wössmann 2003) is also crucial for how socio-economic background affects educational achievement.

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

In this article, we argue that central exams might not only reduce social inequality in school achievement but also mitigate the effect of tracking on inequality. There are two reasons why central exams alleviate the "tracking effect." First, when there are central exit examinations, schools are held accountable and therefore place students in tracks based on objective indicators, not on the basis of social background. Second, central exit exams make schools accountable for *all* tracks, thereby making it likely that relatively more resources are allocated to lower tracks. Whereas some evidence indicates that central exams reduce the effect of tracking on the inequality of educational opportunity (Ayalon and Gamoran 2000), no study to date has empirically tested this hypothesis for a larger number of countries.

Our main objective is to examine the effect of tracking on the inequality of educational opportunity in the context of countries with central exams and those without. The study will focus on secondary education because this is the area in which the institutionalization of curricular tracking has progressed most significantly.[1] Our work begins from the widely supported assumption that student performance is at least partly affected by the organization of educational systems. We consider the effect on students' achievement of curricular tracking between schools and central exit examinations, both independently and in interaction with each other.

To find an answer to our research question, we will analyze data from the OECD's Programme for International Student Assessment (PISA) of 2009. These data allow for sufficient levels of between-country variation in terms of tracking and central exit examinations. We examine data at two levels (students in countries) and at three levels (students in schools in countries). The inclusion of the school level is desirable to obtain more correct estimators of tracking and central examinations because these system characteristics are mediated by the allocation and teaching/learning processes of students in schools. Including the school level is not without problems, such as the possibility of biased coefficients due to over-control or statistical artifacts (Esser forthcoming) or the difference between the total effects of educational system characteristics versus the direct and indirect effects via school characteristics (Dronkers, Van der Velden, and Dunne 2012). Therefore, to demonstrate the difference, we begin without the inclusion of the school level, which is still the state of the art.

Curricular Tracking and Inequality in Learning

An increasing body of internationally standardized data sets has allowed researchers to compare educational outcomes across countries. Empirical findings have shown that in addition to the effects of individuals and schools, educational systems contribute to inequality in educational achievement (for an overview, see Van de Werfhorst and Mijs [2010]). Within different educational systems, curricular tracking—also referred to as "ability grouping," "stratification," "streaming," "sorting," "differentiation," and "placement policies"—is one of the most frequently studied features.

Tracking has been conceptualized as the practice of allocating students into school programs or classes that are homogeneous in terms of "cognitive ability" (Oakes 2005[1985]; Marks 2006; Horn 2009). These practices occur in a variety of ways: students can be allocated to separate schools that offer different curricula; students can be divided into different types of education (or "curricular tracks") within the same school (often labeled academic, vocational, or general); or students can be placed into classes at different levels (or "streams") for the same subject within the same school. Until now, cross-national research has focused primarily on tracking featuring different types of education (either between or within schools) because this is a characteristic that varies more between countries than within school ability groupings (Van de Werfhorst and Mijs 2010). In the present work, we argue that tracking is not a binary variable but rather is gradated, meaning that educational systems can be more or less tracked. The level of tracking depends on several factors, all of which are important, such as the age of first selection, length of the tracked curriculum, and number of tracks.[2]

As mentioned above, tracking occurs not only between but also within schools. It is difficult to see within-school tracking as a variable at the level of *national* systems; in educational systems that have within-school tracking, such as the United States, tracking policies actually differ widely between schools (Oakes 2005[1985]). It could be argued that all educational systems inherently involve some form of tracking. However, the phenomenon is arguably most strongly institutionalized in educational systems with curricular tracking between schools because these societies hold stronger beliefs that the early allocation of students to a large number of tracks is acceptable (LeTendre, Hofer, and Shimizu 2003). In this study, we classify 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 in which tracks exist within schools, are subject-specific, or can be changed between school grades.

One important motivation for increasing the level of tracking is allowing a heterogeneous student population to choose the education that best matches their interests and learning abilities, thereby maximizing the average student achievement (Hanushek and Wössmann 2006). However, this claim is not well substantiated empirically (e.g., Schütz, Ursprung, and Wössmann 2008), and many studies note the detrimental effects of tracking on the performance of less-advantaged students, such as the exacerbation of social inequalities (Hallinan 1994; Van de Werfhorst and Mijs 2010). More specifically, tracking affects the inequality of educational opportunity by two key mechanisms.

The first mechanism is the "social selection mechanism," in which socioeconomic background plays a more significant role in the allocation of students to different tracks, especially when tracking occurs at a younger age. This is explained by the life course hypothesis: one's social background has a greater effect on one's success at a younger age and declines with age (Shavit and Blossfeld 1993). Studies have found that the influence of social background on the choice of school type is greater when educational systems are tracked than in comprehensive systems because students are placed in one of a large number of hierarchical tracks at a young age (Marks 2005; Hanushek and Wössmann 2006).[3]

The second mechanism (or group of mechanisms) argues that tracking affects the inequality of educational opportunity *after* students are selected into different tracks. Relevant factors include socio-psychological aspects, means of instruction, resources available, and peer effects. For example, students who begin in more demanding curricular tracks tend to show greater gains in learning (Marks 2006). An important reason why different curricular tracks lead to varying levels of student achievement is resource endowment, meaning that more resources are allocated to higher tracks, thereby increasing between-track differences in student achievement (Brunello and Checchi 2007). More generally, students in higher curricular tracks tend to benefit from better educational resources (Figlio and Page 2002).

Because of the impact of social selection and the effects of tracking, we hypothesize that the more tracked an educational system is, the more student performance is affected by social background (*hypothesis 1a*).

When studying the relationship between educational systems and the inequality of educational opportunity, it is important to consider school context. Classroom composition and peer interactions may have different effects in tracked and comprehensive systems for students from different social backgrounds (Richer 1976; Hoxby 2000; Wilkinson et al. 2002; Entorf and Lauk 2008). However, including the school level in the analysis changes the interpretation of the tracking effect on inequality of educational opportunity, because in the most tracked educational systems tracking takes place *between* schools. In a strongly tracked educational system, such as that of Germany or the Netherlands, pupils are (often) allocated to different schools and/or locations based on their earlier performance (and based on parental background). In a model that includes the school level, we thus examine the effect of social background on achievement within schools (or, in other words, given the school students attend). These effects are likely to be smaller in more tracked educational systems, as the major source of variation in these systems is between and not within schools.

The main reason why tracked educational systems show a stronger correlation between socio-economic background and achievement is that schools are more important in such systems due to the track placement of students between schools. In tracked systems, pupils with differing social class backgrounds are more strongly segregated between schools, and as a consequence there is less variance in achievement of pupils with differing class backgrounds within schools. Within schools and school types, children are more similar on observed and unobserved characteristics than within schools in untracked systems. Therefore, within schools, students from higher socio-economic backgrounds actually achieve *less* in highly tracked educational systems than their counterparts in comprehensive systems (Dunne 2010; Dronkers, Van der Velden, and Dunne 2012).

Our hypothesis on how tracking affects the relationship between socio-economic background and student achievement thus changes when we take the school level into account. When including the school level in our model, we expect that socio-economic background will have a smaller effect on student performance in educational systems with more tracking (*hypothesis 1b*).

How Central Examinations Interfere

In this article, we are interested mainly in the question of whether the effect of tracking on inequality in educational opportunity changes significantly depending on the existence of central exit examinations.

Central Exit Examinations

Standardization can be defined as “the degree to which the quality of education meets the same standards nationwide” (Allmendinger 1989, 233). One of the most important aspects of standardization is school accountability, the degree to which schools have standardized incentives to perform (Wössmann 2003).

Incentives to increase the accountability of school performance include the monitoring of teacher quality, inspections, and nationally recognized exams. Central examinations are the most vital tool for accountability because 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.

There is empirical support for a substantial positive association between central examinations and student performance (Bishop 1997; Jürges, Schneider, and Büchel 2005; Wössmann 2003). It is argued that this association is due to increased "signaling" of academic achievement (Bishop 1997). Central exams allow students' performance to be compared by future employers and higher-education institutions. This comparison has implications for the rewards and sanctions of learning effort and serves as an incentive for students to perform. Chiang (2009) showed that school accountability improved student performance through an improved allocation of school resources.

The suggested positive effect of central examinations on educational performance has also been disputed. Several studies, mainly investigating the effect of state exams in the United States, show that exit exams do not increase student achievement (Grodsky, Warren, and Kalogrides 2009; Hout and Elliot 2011; see also Hanushek, Warren, and Grodsky [2012]). Opponents of central examinations claim that exams are not an appropriate way to measure performance because the teachers must prepare their students for the tests; therefore, testing leads to unethical test preparation (Jones, Jones, and Hargrove 2003). Moreover, it is argued that "teaching to the test" has negative side effects, leading teachers and students to pay less attention to content that is not tested and to be less engaged with critical thinking (Amrein and Berliner 2003).

Literature on the relationship between central exams and educational inequality is scarce, and the findings are mixed. Although some research suggests an equalizing effect of central exams on student performance (Bishop 1997; Gamoran 1996), other research indicates that the positive effect of central exams on student performance is slightly smaller for low-SES students, indicating an increase in inequality (Wössmann et al. 2009). A third strand finds no relationship between central exams and the association between socioeconomic status and performance (Horn 2009).

The Interaction between Tracking and Central Examinations

How do we expect central examinations to alter the tracking effect? So far, only one study has examined this possible interaction. Ayalon and Gamoran (2000) compared Israel and the United States and found that in the US curriculum tracking led to lower overall achievement and more inequality in performance, whereas the opposite effects were found for Israel. Their explanation was that "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 adopt these authors' suggested explanation and consider the differential impact of tracking in the presence or absence of central exams for a large number of countries. We argue that there are two main explanations for why central exit exams counterbalance the relationship between tracking and inequality of educational opportunity: (1) school accountability and (2) rising lower-track achievement.

First, in countries with central exit examinations, schools are held more directly accountable for their performance. In these systems, it is in the interest of schools for students to graduate at the highest possible level of achievement. Therefore, students are allocated to educational programs according to academic criteria. In the absence of central examinations, in contrast, academic achievement is a less important factor in the allocation of students to tracks, and the secondary effects of parental background 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 in choosing a track. Middle-class children may thus enter a more academic track even if their ability is not quite sufficient. For intermediate ability levels, in particular, this situation may lead to substantial differences in track placement depending on children's social-class background (Boudon 1974; Breen and Yaish 2006). However, when schools are held accountable by means of central examinations, plausibly more weight is placed on students' ability levels rather than their social background. Our argument is not that selection in scholastic achievement makes track placement fully equal across social classes. Rather, inasmuch as social-class differences in preferences, ambitions, and forms of stimulation are independent of student performance, the impact of such factors is reduced in systems with high accountability.

Second, central exams are 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. Thus, schools and teachers attempt to maximize the performance of students in all tracks. Therefore, one outcome of central exams may be that schools redistribute resources so that students in lower tracks benefit as well, and schools may be motivated to hire better teachers for all tracks. In countries with central examinations, schools are likely to invest relatively more in lower-track students, whereas this is not the case in comprehensive educational systems. We argue that central exams compensate for the post-placement track effects that are usually beneficial only for students in higher tracks (cf. Brunello and Checchi 2007).

Briefly, our proposition is that central exams attenuate the effect of tracking on the inequality of educational opportunity by (1) decreasing social selection and (2) reducing the inequalities that arise after track placement. Thus, we expect the association between tracking and the inequality of educational opportunity to be weaker when central examinations are present (*hypothesis* 2).

Data and Methods

Data

The main source of data for this study is the Programme of International School Assessment (PISA) for the year 2009. This program, developed by the OECD, is an international study that 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 $N_1 = 261{,}578$ (36 countries). We follow the data analysis manual of the PISA data (OECD 2009) and use sampling weights for all our analyses. The sum of weights is equalized for each country such 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 greatest challenge in comparative research into the relationship between educational systems and student achievement involves unobserved heterogeneity at the 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). All between-country heterogeneity in student achievement is included in these fixed effects. Although we focus on central examinations and tracking, there may be several other reasons why students perform differently across countries; for example, countries may invest more in education in general. These cross-national differences are accounted for by the fixed effects. Country fixed effects can be included if the focus is on cross-level interaction terms (as is the case in the present study between educational institutional variables and social background) because the main effects of country-level variables cannot be included in the model. The cross-level interaction effects estimate potential non-linearities in the effects of individual-level variables, which is precisely in line with what is hypothesized in this article.

Fixing the country-level effects has a number of advantages over "normal" multilevel regression models. First, given the non-random sample of countries, multilevel models that include a country level are likely to violate the assumption of normality of the country-level residuals. Second, the sample size at the country level is small (our sample has 36 countries). Consequently, the number of degrees of freedom is limited at the country level.

The general equation for the country fixed effect model is the following:

$$\begin{aligned} Y_i = {} & \beta_{0i} + \beta_1 \text{fem}_i + \beta_2 \text{age}_i + \beta_3 \text{grade}_i + \beta_4 \text{immig1}_i + \beta_5 \text{immig2}_i + \beta_6 \text{ESCS}_i \\ & + \sum_{k=k-1} \beta_x D_k + \beta_8 \text{ESCS*track}_i + \beta_9 \text{ESCS*CE}_i + \beta_{10} \text{ESCS*track*CE}_i + e_i . \quad (1) \end{aligned}$$

In equation 1, Y_i is the score on the mathematics 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 (dummy 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_i is the error term.

If, as our hypotheses predict, tracking generally decreases equality of educational opportunity but does so less in systems with central examinations, β_8 should be positive and β_{10} negative. We use a three-way interaction to investigate hypothesis 2, as we expect that the two-way interaction effect of tracking and socio-economic background on student achievement is dependent on the presence or absence of central exams. Although three-way interactions are not easy to interpret, in this case it does provide the true test of our hypothesis, as the three-way interaction shows us if there is a significant difference in the tracking effect on the equality of educational opportunity between systems with and without central exit exams.

In addition, the PISA data allow us to incorporate the school level into our analysis ($N_2 = 9{,}834$ schools). By using multilevel models in which we take schools into account, we can 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*), and countries are again added as a fixed effect. The general equation for these models is

$$\begin{aligned} Y_{ij} = {} & \beta_{0ij} + \beta_1 \text{fem}_{ij} + \beta_2 \text{age}_{ij} + \beta_3 \text{grade}_{ij} + \beta_4 \text{immig1}_{ij} + \beta_5 \text{immig2}_{ij} \\ & + \beta_6 \text{ESCS}_{ij} + \sum_{k=k-1} \beta_x D_k + \beta_8 \text{ESCS*track}_{ij} + \beta_9 \text{ESCS*CE}_{ij} \\ & + \beta_{10} \text{ESCS*track*CE}_{ij} + e_{ij} + u_{0j}. \end{aligned} \tag{2}$$

Equation 2 is very 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).

Variables

For the dependent variable, we use student performance in mathematics. The PISA 2009 data set includes three cognitive variables: mathematics, reading, and science. We use mathematics because it is most clearly learned at school rather than at home (e.g., Coleman 1975; Scheerens and Bosker 1997). It is important to note that using either reading or science as dependent variables provides evidence that is even stronger than the results we will show for mathematics as the dependent variable. A brief summary of our results using science and reading as dependent variables can be found in appendices A1 and A2 (supplementary material online), respectively. The mathematics score is scaled so that the OECD mean score is 500 with a standard deviation of 100 (OECD 2012).

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 2009, 96). The OECD provides five plausible values for each domain. Although these five items are highly correlated, averaging plausible values at the individual level leads to biased estimates (OECD 2009, 100). We therefore estimate all models separately for each plausible value, average all parameters between the five models, and calculate standard errors to account for the variance both within and between plausible values (OECD 2009).

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 2012). In addition to socio-economic status, we control our results for the student’s sex (female = 1) and age, and include a variable for immigration status (first generation, second generation, and native). We finally add the student’s grade, which is essential because PISA interviews 15-year-olds independent of grade level.[4]

At 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 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. The descriptive statistics for all the individual- and school-level variables used in the total sample can be found in table 1.

The country-level data are obtained from a macro-level data set that was composed of a variety of sources by Bol and Van de Werfhorst (2013a, 2013b).[5] Our main interest 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 are not affected in response to student achievement.

The index of tracking that we use is a combined index of three measures that indicate the level of tracking in secondary education. First, the age of first selection (OECD 2006, 102, 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 indicates how many tracks there are for the respondents in the PISA sample (OECD 2006, 102, table A7.1). The final indicator we use to operationalize tracking is how long the tracked curriculum lasts. This final indicator calculates the percentage of the total compulsory curriculum that is tracked and is calculated on the basis of data

Table 1. Descriptive Statistics for All Used Variables

	Mean	SD	Minimum score	Maximum score
Individual-level variables				
Plausible values				
Plausible value mathematics 1	501.939	93.803	42.460	901.860
Plausible value mathematics 2	502.031	93.904	34.830	900.460
Plausible value mathematics 3	501.990	93.860	4.450	908.870
Plausible value mathematics 4	501.956	93.951	3.670	862.290
Plausible value mathematics 5	501.946	93.934	5.230	953.270
Female	0.502	0.500	0.000	1.000
Grade	9.679	0.703	7.000	13.000
Age	15.776	0.291	15.170	16.330
Immigrant status				
Native citizen	0.889	0.315	0.000	1.000
2nd-generation immigrant	0.059	0.235	0.000	1.000
1st-generation immigrant	0.053	0.223	0.000	1.000
ESCS index	0.034	0.955	–6.037	3.408
School-level variables				
Public school	0.820	0.384	0.000	1.000
Student/teacher ratio	12.614	4.952	0.565	109.500
School location				
Village	0.083	0.275	0.000	1.000
Small town	0.229	0.420	0.000	1.000
Town	0.343	0.475	0.000	1.000
City	0.228	0.420	0.000	1.000
Large city	0.118	0.322	0.000	1.000
Mean school ESCS	0.031	0.562	–3.544	1.724

Note: The descriptive statistics for the individual-level variables are calculated using our empirical sample of PISA 2006 ($N = 261{,}578$). The descriptive statistics for the school-level variables are calculated by using only those observations that had no missings for either of the school variables ($N = 239{,}468$).

from OECD (Brunello and Checchi 2007, 799). We perform a factor analysis on these three indicators and save the results as regression coefficients.[6] Countries with an educational system that has an average level of tracking score 0 on this index. A negative deviation signals less tracking than average, whereas a positive deviation signals more tracking. It is important to note that the factor analysis refers to a larger sample of countries than is studied in this article. Because a factor analysis estimates the relative position of all observations, as many countries as possible are included (in Bol and Van de Werfhorst [2013a]).

The criteria for the central examinations indicator are based on the operationalization of Bishop (1997, 260), who proposes five criteria for central exams: (1) the diploma has real consequences and is not 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) the exam is part of secondary education and covers most of the student population in secondary education. The data for central examinations are derived from the *European Glossary on Education*, particularly from the section on examinations, qualifications, and titles (Eurydice 2004). This report provides a summary of all examinations in a country and indicates whether the exams are enforced by a national institution. If countries meet both criteria, they are

Figure 1. The tracking index and central exit examinations across countries

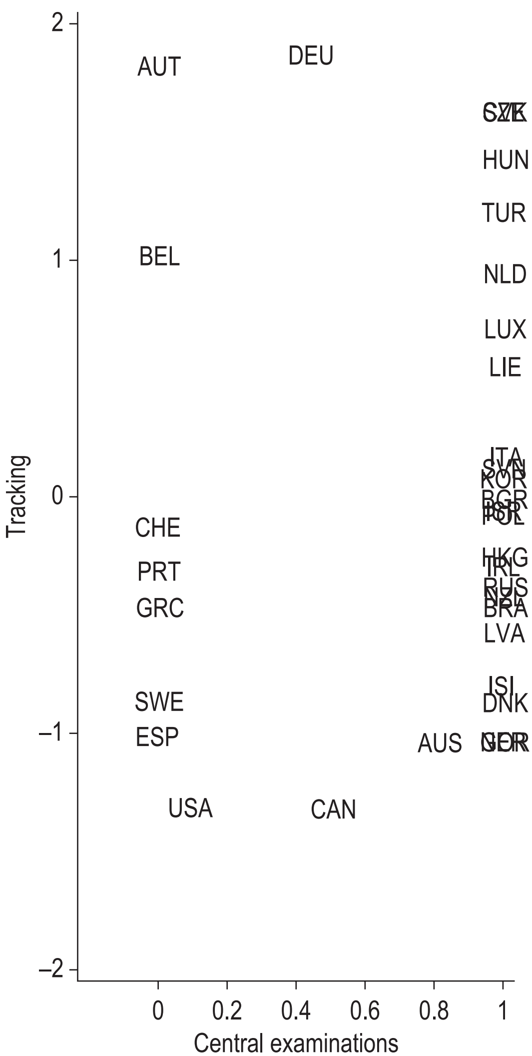


Table 2. Descriptive Statistics for Country-Level Variables

	Tracking index	Central exams	Mean ESCS	SD ESCS
Australia	–1.043	0.81	0.320	0.757
Austria	1.817	0	0.087	0.825
Belgium	1.018	0	0.233	0.912
Bulgaria	–0.019	1	–0.099	0.982
Canada	–1.321	0.51	0.466	0.821
Czech Rep.	1.621	1	0.017	0.742
Denmark	–0.870	1	0.143	0.935
Finland	–0.870	1	0.412	0.780
France	–0.474	1	–0.115	0.841
Germany	1.862	0.44	0.198	0.900
Great Britain	–1.043	1	0.184	0.781
Greece	–0.474	0	0.029	0.986
Hong Kong	–0.273	1	–0.809	1.012
Hungary	1.421	1	–0.154	0.946
Iceland	–0.805	1	0.715	0.884
Ireland	–0.302	1	0.061	0.847
Israel	–0.063	1	–0.011	0.880
Italy	0.166	1	–0.096	0.978
Japan	–0.474	1	–0.008	0.722
Korea Rep.	0.072	1	–0.129	0.820
Latvia	–0.576	1	–0.048	0.859
Liechtenstein	0.546	1	0.083	0.941
Luxembourg	0.700	1	0.222	1.090
Netherlands	0.937	1	0.314	0.845
New Zealand	–0.419	1	0.098	0.779
Norway	–1.043	1	0.476	0.735
Poland	–0.083	1	–0.215	0.914
Portugal	–0.327	0	–0.239	1.159
Russia	–0.386	1	–0.161	0.794
Slovakia	1.621	1	–0.088	0.837
Slovenia	0.117	1	–0.057	0.877
Spain	–1.020	0	–0.249	1.055
Sweden	–0.870	0	0.338	0.811
Switzerland	–0.138	0	0.026	0.856
Turkey	1.201	1	–1.151	1.204
United States	–1.321	0.09	0.156	0.920

coded 1. Because the report covers only European countries, our data are supplemented and cross-checked with data from Fuchs and Wössmann (2007, 438) and Wössmann et al. (2009, 123). The vast majority of countries receive a value of 0 (no exams) or 1 (exams). However, following Wössmann et al. (2009), four countries receive a different value (Australia, Canada, Germany, and the United States) because central examinations are held in only some states or provinces within the country.

In figure 1, the countries are displayed by their level of tracking and the existence of central examinations. It is evident that there are no "empty cells" in our institutional mix; there are countries with and without central exams that vary in the extent to which they have tracked educational systems.

Table 2 shows the scores for all countries for both institutional indicators that are relevant for the present analysis. Furthermore, table 2 contains the mean and standard deviation in the ESCS variable to indicate the amount of inequality in test scores in each country. We observe that there is considerable variation in the mean level of the socio-economic status of the students, which is lowest in Turkey and highest in Iceland. The final column shows the standard deviation and the level of dispersion of test scores. Here, we see that countries are more alike in this aspect, and the standard deviation around the mean of most countries is close to 1.

Results

We will first discuss the models that examine individuals nested in countries without considering the school level (following equation 1). We then will show the results of models in which 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 nearly 8 percent of the total variation in students' scores on the PISA mathematics test. In model 1, the relationship 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 expected direction. 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, 14.3 points less than boys, whereas first- and second-generation immigrants score significantly worse (12.9 and 10.1 points, respectively) than native students. Grade year is positively associated with performance; students 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 33.3 points.

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 because

Table 3. Regression Models with Country Fixed Effects (dependent variable: performance on PISA mathematics test)

	Model 0	Model 1	Model 2	Model 3	Model 4
Individual level					
Country fixed effects	*yes*	*yes*	*yes*	*yes*	*yes*
Female		–14.267**	–14.259**	–14.270**	–14.272**
		(0.461)	(0.461)	(0.462)	(0.461)
Grade		46.946**	46.807**	46.936**	46.972**
		(0.516)	(0.518)	(0.518)	(0.518)
Age		–9.465**	–9.412**	–9.462**	–9.477**
		(0.833)	(0.833)	(0.834)	(0.833)
Native		ref.	ref.	ref.	ref.
2nd–generation immigrant		–12.938**	–12.769**	–12.915**	–12.836**
		(1.067)	(1.069)	(1.072)	(1.072)
1st-generation immigrant		–10.125**	–10.105**	–10.248**	–10.197**
		(1.447)	(1.446)	(1.446)	(1.447)
Socio-economic status (ESCS index)		33.282**	33.279**	30.908**	31.181**
		(0.284)	(0.284)	(0.421)	(0.427)
Interactions					
ESCS*tracking			1.757**	1.544**	3.139**
			(0.264)	(0.260)	(0.429)
ESCS*central exams				3.275**	3.084**
				(0.559)	(0.568)
ESCS*central exams*tracking					–2.545**
					(0.607)
Constant	517.043**	191.384**	192.553**	191.905**	191.564**
	(0.904)	(12.715)	(12.722)	(12.725)	(12.723)
R^2	0.079	0.258	0.258	0.258	0.259
Observations	261,578	261,578	261,578	261,578	261,578

Source: PISA 2009, own calculations for 36 countries (see table 3).
Note: Standard error in brackets. All standard errors calculated by taking into account the variance both between and within plausible values. Sampling weights were used in all analyses.
** $p < .01$, two-tailed tests

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 one-point increase on the scale of tracking, the predicted effect of socio-economic status on performance in the mathematics test increases by 1.8 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 positive, indicating that in countries with central examinations, there is greater inequality in student performance across different socio-economic backgrounds. This finding may be explained by the increasing competition between schools when accountability systems are established (Hanushek and Raymond 2005). However, we do not attach too much significance to this finding because in the later models (see tables 4 and 5) we show that the positive interaction between social class and central examinations is not robust once the school level is included.

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 an effect of –2.5, which is highly significant. The negative effect of the three-way interaction shows that the relationship between tracking and inequality in student performance by socio-economic background is attenuated when central exams are applied.

If we compare the interaction between tracking and socio-economic background in the different models, we see that this effect for model 4 is nearly double that of models 2 and 3. For each one-point increase on the tracking index, the effect of ESCS on student performance increases by 3.1 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 relationship between socio-economic background and student performance becomes even larger than previously thought. However, when central exams are implemented, this negative effect of tracking is offset. This finding is shown best by an example.

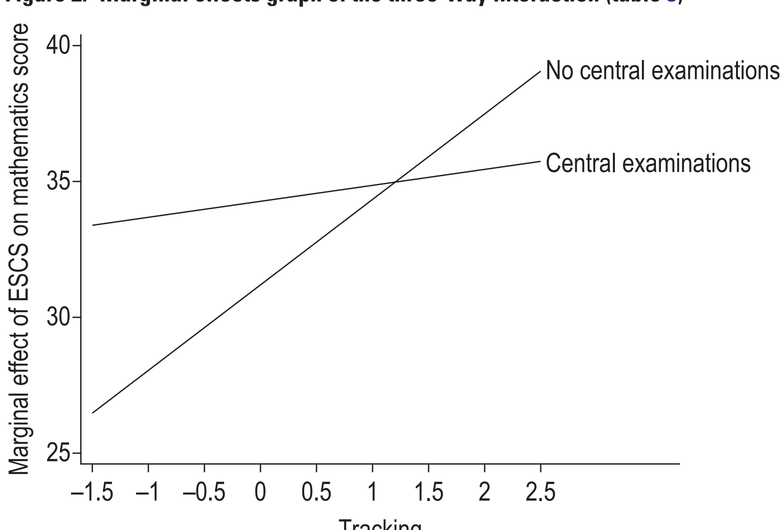
Following the lead of Aiken and West (1991), we can estimate the marginal effect P of ESCS on performance for varying scores of tracking and central exams using the following equation:

$$P = 31.2 + \left(3.1 \times \text{tracking}\right) + \left(3.1 \times \text{CE}\right) + \left(-2.5 \times \text{tracking} \times \text{CE}\right). \qquad (3)$$

This formula shows that the positive effect of tracking is largely offset by the implementation of central exams. In a highly tracked country (measuring 2 on the tracking index), the predicted marginal effect of ESCS on school performance would be 35.5 rather than 37.4, a difference of nearly two 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 attenuated, although tracking still increases the inequality of educational opportunity in systems both with and without central exams. Therefore, this finding confirms hypothesis 2.

Figure 2 depicts the above calculation and the main finding of table 3. The graph shows the marginal effect of ESCS on the mathematics score for the changing values of our tracking index. We plotted two lines: one in the absence and one in the presence of central examinations. It is clear that in the absence of central examinations, educational institutions with tracking are associated with larger social inequalities in learning. Figure 2 also makes clear that the predicted effect of ESCS on the math test score is lowest in the absence of both central exams and tracking. In a country with a comprehensive educational system (measuring –1 on the tracking scale) and no central exams, the predicted effect of ESCS is only 28.1. Our major contribution to existing literature is, however, that the detrimental effects of tracking on the equality of educational opportunity are larger in the absence of central exams.

Figure 2. Marginal effects graph of the three-way interaction (table 3)



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.

Adding the School Level to the Analysis

As discussed in an earlier section of this study, it is important to consider the school level. However, this consideration also changes the interpretation of the country-level effect of tracking (see hypothesis 1b). Schools play an intermediary role because they are located between educational systems and students' opportunities for educational performance. In many countries, the tracking effect is largely caused by the allocation of students to schools of different educational levels. We are predominantly interested in determining whether our main result—the negative three-way interaction between tracking, central exams, and socio-economic background—is also confirmed in a model that 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 added to our analysis. In the null model of table 4, we can see that schools make up 35.2 percent of the total variance in student performance on the PISA mathematics 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 perform even worse (–19.3) than was estimated in the same model in table 3. Moreover, the effect of ESCS on student performance decreases to approximately two-thirds of the original estimate when we include the school level and thus take into account that students are not randomly clustered across schools.

When we consider the interaction effects between socio-economic background and our contextual variables, even more important implications arise. 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 that 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 mathematics test decreases by 4.9 points. We therefore confirm our hypothesis 1b.

If we take the schools into account by adding a random intercept, 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 has 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 based on parental background). It is important to note that although the interaction effect of tracking and socio-economic background is negative in our three-level models, we are not arguing that tracking decreases inequality in student achievement along socio-economic lines. Rather, our results show that the effect of tracking on the relationship between social class and student performance is due mainly to the placement of students in different schools (cf. Dronkers, Van der Velden, and Dunne 2012).

Model 3 in table 4 also shows that the positive interaction effect between ESCS and central examinations that we found in table 3 disappears when we

Table 4. Multilevel Linear Regression Models with a Random School Intercept and Fixed Country Effects (dependent variable: performance on PISA mathematics test)

	Model 0	Model 1	Model 2	Model 3	Model 4
Individual level					
Country fixed effects	*yes*	*yes*	*yes*	*yes*	*yes*
Female		–19.303**	–19.435**	–19.435**	–19.439**
		(0.472)	(0.473)	(0.473)	(0.473)
Grade		43.175**	43.207**	43.212**	43.235**
		(0.660)	(0.659)	(0.660)	(0.660)
Age		–8.881**	–8.843**	–8.845**	–8.855**
		(0.759)	(0.758)	(0.758)	(0.758)
Native		ref.	ref.	ref.	ref.
2nd-generation immigrant		–13.080**	–13.526**	–13.532**	–13.446**
		(1.083)	(1.072)	(1.075)	(1.074)
1st-generation immigrant		–11.316**	–11.379**	–11.385**	–11.400**
		(1.373)	(1.376)	(1.379)	(1.379)
Socio-economic status (ESCS index)		18.918**	18.762**	18.640**	18.932**
		(0.342)	(0.325)	(0.436)	(0.434)
Interactions					
ESCS*tracking			–4.944**	–4.954**	–3.432**
			(0.297)	(0.292)	(0.429)
ESCS*central exams				0.169	–0.082
				(0.626)	(0.631)
ESCS*central exams*tracking					–2.407**
					(0.666)
Constant	511.178**	223.618**	221.221**	221.198**	220.943**
	(2.475)	(12.705)	(12.747)	(12.739)	(12.753)
Σu (school)	2861.274**	1938.437**	1971.150**	1971.084**	1970.040**
	(26.542)	(19.415)	(20.013)	(20.011)	(20.052)
Σe	5263.017**	4624.682**	4609.979**	4609.979**	4609.526**
	(17.844)	(15.554)	(15.722)	(15.722)	(15.748)
ICC (school)	0.352	0.296	0.300	0.300	0.299
–2LL	3,006,037	2,970,119	2,969,444	2,969,443	2,969,414
Observations	261,578	261,578	261,578	261,578	261,578
Number of schools	9,834	9,834	9,834	9,834	9,834

Source: PISA 2009, own calculations for 36 countries (see table 3).
Note: Standard error in brackets. All standard errors calculated by taking into account the variance both between and within plausible values. Sampling weights were used in all analyses.
** $p < .01$, two-tailed tests

add the school level.[7] Most importantly, we again confirm the main finding: the three-way interaction between socio-economic background, tracking, and central examinations is negative and significant. Even when the role of schools is taken into consideration, in systems without central examinations, tracking is still more detrimental to the inequality of educational opportunity than it is in systems with central examinations.

Adding the school level also allows us to control our findings for school-level characteristics (table 5). Public 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 and Robert 2008). However, model 8 also shows that when we take the socio-economic composition of schools into account, this effect changes direction. Therefore, the negative effect of public schools is due largely to the composition of these schools.

The student/teacher ratio is positively related to student performance, suggesting that students in schools with larger 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, all of which show significant effects. In this model, we can also see that both the two-way interaction between tracking and ESCS and the three-way interaction provide 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. In model 8, we see that, unsurprisingly, the social composition of schools has a large effect on student performance (60.2).

Most importantly, 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 associated with a smaller effect of tracking on the relationship between socio-economic background and educational achievement. When we again calculate the marginal effect P using the results from model 8 (see formula below), we find that in a highly tracked country (measuring 2 on the tracking index), the predicted marginal effect of ESCS on school performance is 6.3 with and 11.3 without central exit examinations, a difference of five points:

$$P = 17.1 + \left(-2.9 \times \text{tracking}\right) + \left(0.4 \times \text{CE}\right) + \left(-2.7 \times \text{tracking} \times \text{CE}\right). \quad (4)$$

Figure 3 depicts this finding graphically. It shows the marginal effect of ESCS for changing values of tracking and central exams, this time based on table 5. The figure shows that the association between tracking and the marginal effect of ESCS on student performance is smaller in educational systems that have central exams.

Table 5. Multilevel Linear Regression Models with School-Level Controls (dependent variable: performance on PISA mathematics test)

	Model 5	Model 6	Model 7	Model 8
Individual level				
Country fixed effects	*yes*	*yes*	*yes*	*yes*
Female	–19.129**	–19.152**	–19.169**	–19.304**
	(0.486)	(0.487)	(0.487)	(0.486)
Grade	42.869**	43.356**	43.326**	42.278**
	(0.668)	(0.686)	(0.687)	(0.684)
Age	–8.716**	–8.842**	–8.862**	–8.462**
	(0.774)	(0.799)	(0.800)	(0.799)
Native	ref.	ref.	ref.	ref.
2nd-generation immigrant	–12.918**	–12.921**	–13.068**	–12.712**
	(1.110)	(1.147)	(1.152)	(1.145)
1st-generation immigrant	–11.308**	–10.424**	–10.550**	–10.277**
	(1.401)	(1.446)	(1.453)	(1.449)
Socio-economic status (ESCS index)	18.924**	18.788**	18.743**	17.095**
	(0.439)	(0.452)	(0.452)	(0.457)
Interactions				
ESCS*tracking	–3.356**	–3.378**	–3.360**	–2.936**
	(0.435)	(0.448)	(0.448)	(0.451)
ESCS*central exams	0.011	–0.073	–0.067	0.386
	(0.644)	(0.665)	(0.666)	(0.662)
ESCS*central exams*tracking	–2.468**	–2.302**	–2.307**	–2.694**
	(0.675)	(0.694)	(0.695)	(0.694)
School-level				
Public school	–13.227**	–12.419**	–11.477**	7.932**
	(1.609)	(1.699)	(1.707)	(1.463)
Student/teacher ratio		0.624**	0.507**	–0.095
		(0.161)	(0.159)	(0.116)
School located in village			ref.	ref.
School located in small town			6.173**	–3.716*
			(1.742)	(1.643)
School located in town			9.228**	–7.023**
			(1.757)	(1.634)
School located in city			12.332**	–11.344**
			(1.884)	(1.724)

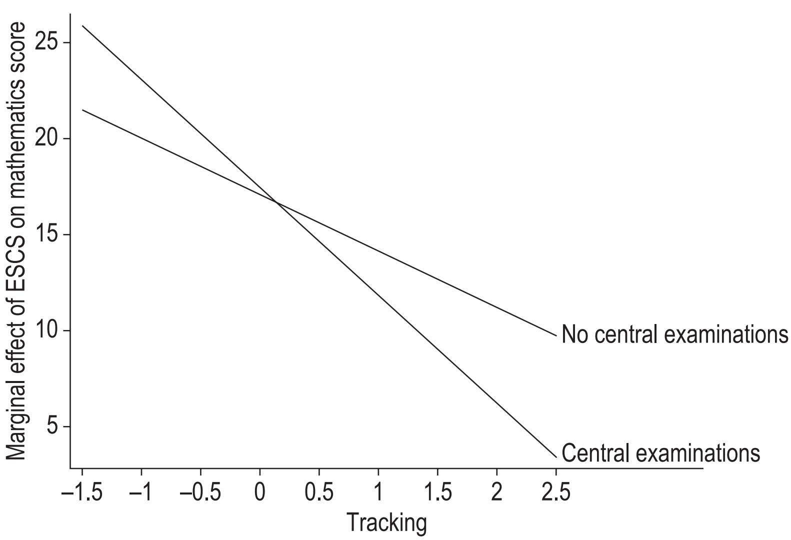
(Continued)

Table 5. *Continued*

	Model 5	Model 6	Model 7	Model 8
School located in big city			12.937** (2.497)	–10.999** (2.122)
Mean school ESCS				60.244** (1.258)
Constant	230.424** (13.117)	218.906** (13.833)	210.024** (13.866)	211.831** (13.732)
Σu (school)	1922.283** (20.188)	1928.711** (20.976)	1913.861** (20.747)	1229.355** (13.572)
Σe	4638.467** (16.265)	4622.317** (16.834)	4621.666** (16.855)	4622.199** (16.941)
ICC (school)	0.293	0.294	0.292	0.209
–2LL	2,852,957	2,692,882	2,689,761	2,686,711
Observations	253,921	239,865	239,468	239,468
Number of schools	9,515	8,972	8,959	8,959

Source: PISA 2009, own calculations for 35 countries (see table 3).
Note: Standard error in brackets. All standard errors calculated by taking into account the variance both 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 several countries and the complete absence of school data for France.
** $p < .01$, * $p < .05$, two-tailed tests

Figure 3. Marginal effects graph of the three-way interaction (table 5)



Note: Based on model 4, table 5. 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.

Conclusion

Many studies have shown that greater social inequality is found in the achievement levels of students in educational systems where students are separated early into different curricular tracks than in comprehensive systems (Brunello and Checchi 2007; Van de Werfhorst and Mijs 2010; Marks 2005). However, such studies have often ignored the question of whether the relationship between tracking and 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 relationship has not previously been explored. In this study, we sought to fill this gap and studied one aspect of educational systems that has a profound influence on how tracking is related to inequality: central examinations.

Using data from the PISA 2009, we confirm earlier findings that inequality in learning due to students' socio-economic background is larger in educational systems that are strongly tracked 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 affected by whether 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. It is important to note that our study does not argue that central exams increase the mean performance of all students. What we argue is that the detrimental effect of tracking on the relationship between social background and student achievement is attenuated in the presence of central exit exams.

The inclusion of the school level in our analysis demonstrates that the measured and unmeasured characteristics of schools also affect the impact of tracking on the importance of social background for learning. Differences between schools that are a result of between-school tracking—selection, track, and socio-economic composition effects—are thus accounted for. Even in the analyses that incorporated the school level, a negative three-way interaction between socio-economic background, tracking, and central exams was found.

We have two explanations for our main finding that the interaction effect between tracking and socio-economic background on achievement in mathematics is attenuated in systems with central exit examinations.

First, schools are more easily held accountable for their performance if they must participate in central examinations (cf. Wössmann et al. 2009). Selection into tracks therefore occurs more on the basis of objective indicators of ability than socio-economic background. So-called "secondary effects" are likely to be smaller when a country implements central examinations.

Second, central exams are likely to particularly affect schools that provide lower (pre)vocational tracks, which include a higher proportion of children from lower socio-economic backgrounds. Therefore, children from disadvantaged backgrounds benefit more from the educational system than children from higher socio-economic classes, who benefit more from their family resources. In countries with centralized examinations, therefore, the lower tracks are less evidently considered “waste bins for the untalented” than they are in countries where schools are less easily held accountable for performance.

These findings are relevant for educational policy because they shed new light on the relationship between tracking and educational inequality. The OECD (2007, 14) has suggested that “limiting early tracking and streaming and postponing academic selection” is an important step to promoting equity in education. However, our article shows that the negative impact of tracking can be attenuated 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. In addition to the move to reform educational systems toward more comprehensive education, another means of combating inequality in educational opportunity is to incorporate central examinations at various times before and during the secondary-school years. Although we find that tracking still increases social inequalities, the extent to which it does so is lower in countries with central examinations.

Notes

1. Secondary education means education that follows the basic programs of the primary level and includes the final stage of compulsory education. Often, subjects are taught by specialized teachers who conduct mainly classes in their field of specialization. Typically, cross-national research has examined tracking mostly in secondary education (e.g., Marks 2005; Schütz, Ursprung, and Wössmann 2008) because tracking in higher education is less institutionalized (but see Shavit, Arum, and Gamoran 2007). Given that we are studying secondary-school achievement, we 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). Although we acknowledge that track mobility may offset some of the negative effects of tracking, data limitations do not allow us to incorporate this dimension into our analysis.
3. In addition to social background, geographical location can influence a student’s access to school and hence the allocation of students. Unfortunately, our data do not allow us to determine the extent to which between-school homogeneity or heterogeneity is a result of residential segregation.
4. Unfortunately, the PISA data do not provide information on the track placement of students, nor do we have information on the achievement or track placement of students before the age of 15.
5. We use an updated version of the country data that are available from www.thijsbol.com.
6. One factor was extracted from the data, which has an eigenvalue of 1.99.
7. An explanation for this disappearance is the inclusion of the school level, which lowers the coefficients for individual ESCS because part of the variance is now situated at the school level instead of the individual level.

About the Authors

Thijs Bol is assistant professor of sociology at the University of Amsterdam, and member of the Amsterdam Centre for Inequality Studies. His current research focuses on how educational systems affect educational inequalities and the school-to-work transition, the strength of linkages between educational degrees and occupations, and the role of occupations (and occupational closure) in generating wage inequality. Recent publications have appeared in *Research in Social Stratification and Mobility*, *Comparative Education Review*, *Social Science Research*, and *Sociological Spectrum*.

Jacqueline Witschge is a PhD candidate in sociology at the Amsterdam Institute for Social Science Research (AISSR) and the Amsterdam Centre for Inequality Studies (AMCIS). Her research deals with the question of to what degree educational systems, including the organization of civic education, affect the distribution of civic competences across socioeconomic groups. Her main research interests are social stratification and mobility, educational inequality, political socialization, and civic engagement.

Herman G. Van de Werfhorst is professor of sociology at the University of Amsterdam, and director of the Amsterdam Centre for Inequality Studies (AMCIS). He studies the impact of educational institutions on various inequalities. His recent publications on this topic have appeared in the *Annual Review of Sociology, Social Science Research, Comparative Education Review, Demography, Acta Sociologica, European Sociological Review, Research in Social Stratification and Mobility*, *Changing Inequalities and Societal Impacts in Rich Countries: Analytical and Comparative Perspectives*, and Oxford University Press.

Jaap Dronkers holds the chair of international comparative research on educational performance and social inequality at Maastricht University. He publishes on unequal educational attainment, changes in educational opportunities, differences between public and religious schools, achievement of migrants from different origins and destinations, effects of parental divorce, cross-national differences in divorce, and European nobility. He also participates in public debates on topics related to his research. Recent publications have appeared in *International Migration, European Union Politics, European Educational Research Journal, Social Forces,* and *Educational Research and Evaluation*.

Supplementary Material

Supplementary material is available at *Social Forces* online, http://sf.oxfordjournals.org/

References

Aiken, Leona S., and Stephen G. West. 1991. *Multiple Regression: Testing and Interpreting Interactions*. Thousand Oaks, CA: Sage.

Allmendinger, Jutta. 1989. "Educational Systems and Labor Market Outcomes." *European Sociological Review* 5(3):231–50.

Amrein, Audrey L., and David C. Berliner. 2003. "The Effects of High-Stakes Testing on Student Motivation and Learning: A Research Report." *Educational Leadership* 60(5):32–38.

Ayalon, Hanna, and Adam Gamoran. 2000. "Stratification in Academic Secondary Programs and Educational Inequality in Israel and the United States." *Comparative Education Review* 44(1):54–80.

Bishop, John H. 1997. "The Effect of National Standards and Curriculum-Based Exams on Achievement." *American Economic Review* 87(2):260–64.

Bol, Thijs, and Herman G. Van de Werfhorst. 2013a. "The Measurement of Tracking, Vocational Orientation, and Standardization of Educational Systems: a Comparative Approach." *Gini Discussion Paper 81*.

_______. 2013b. "Educational Systems and the Trade-Off Between Labor Market Allocation and Equality of Educational Opportunity." *Comparative Education Review* 57(2):285–308.

Boudon, Raymond. 1974. *Education, Opportunity, and Social Inequality: Changing Prospects in Western Society*. New York: Wiley-Interscience.

Breen, Richard, and John H. Goldthorpe. 1997. "Explaining Educational Differentials: Towards a Formal Rational Action Theory." *Rationality and Society* 9(3):275–305.

Breen, Richard, and Meir Yaish. 2006. "Testing the Breen-Goldthorpe Model of Educational Decision Making." In *Mobility and Inequality: Frontiers of Research in Sociology and Economics,* edited by S. L. Morgan, D. B. Grusky, and G. S. Fields, 232–58. Palo Alto, CA: Stanford University Press.

Brunello, Giorgio, and Daniele Checchi. 2007. "Does School Tracking Affect Inequality of Opportunity? New International Evidence." *Economic Policy* 22(52):781–861.

Chiang, Hanley. 2009. "How Accountability Pressure on Failing Schools Affects Student Achievement." *Journal of Public Economics* 93(9–10):1045–57.

Coleman, James S. 1975. "Methods and Results in the IEA Studies of the Effects of School on Learning." *Review of Educational Research* 45(3):335–86.

Coleman, James S., Ernest Q. Campbell, Carol J. Hobson, James McPartland, Alexander M. Mood, Frederic D. Weinfeld, and Robert L. York. 1966. *Equality of Educational Opportunity* . Washington, DC: US Government Printing Office.

Dronkers, Jaap, and Peter Robert. 2008. "Differences in Scholastic Achievement of Public, Private Government-Dependent, and Private Independent Schools: A Cross-National Analysis." *Educational Policy* 22:541–77.

Dronkers, Jaap, Rolf Van der Velden, and Allison Dunne. 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 diss., Department of Political and Social Sciences, European University Institute.

Duru-Bellat, Marie, and Annick Kieffer. 2000. "Inequalities in Educational Opportunities in France: Educational Expansion, Democratization or Shifting Barriers?" *Journal of Education Policy* 15(3):333–52.

Entorf, Horst, and Martina Lauk. 2008. "Peer Effects, Social Multipliers, and Migrants at School: An International Comparison." *Journal of Ethnic and Migration Studies* 34(4):633–54.

Erikson, Robert, and Jan O. Jonsson. 1996. *Can Education Be Equalized? The Swedish Case in Comparative Perspective*. Boulder, CO: Westview Press.

Esser, Harmut. Forthcoming. "Bildungssysteme und (ethnische) Bildungsungleichheit." *Mannheim Centre for European Social Research, University of Mannheim*, 2014. Unpublished paper.

Eurydice. 2004. *European Glossary on Education. Volume 1, 2nd ed. Examinations, Qualifications, and Titles*. Brussels: Eurydice.

Figlio, David N., and Marianne E. Page. 2002. "School Choice and the Distributional Effects of Ability Tracking: Does Separation Increase Inequality?" *Journal of Urban Economics* 51(3):497–514.

Fuchs, Thomas, and Ludger Wössmann. 2007. "What Accounts for International Differences in Student Performance? A Re-Examination Using PISA Data." *Empirical Economics* 32:433–64.

Gamoran, Adam. 1996. "Curriculum Standardization and Equality of Opportunity in Scottish Secondary Education: 1984–90." *Sociology of Education* 69(1):1–21.

Gamoran, Adam, and Matthew Weinstein. 1998. "Differentiation and Opportunity in Restructured Schools." *American Journal of Education* 106(3):385–415.

Grodsky, Eric, John R. Warren, and Demetra Kalogrides. 2009. "State High School Exit Examinations and NAEP Long-Term Trends in Reading and Mathematics, 1971–2004." *Educational Policy* 23(4):589–614.

Hallinan, Maureen T. 1994. "Tracking: From Theory to Practice." *Sociology of Education* 67(2):79–84.

Hanushek, Eric A., and Margaret E. Raymond. 2005. "Does School Accountability Lead to Improved Student Performance?" *Journal of Policy Analysis and Management* 24(2):297–327.

Hanushek, Eric A., John R. Warren, and Eric Grodsky. 2012. "Evidence, Methodology, Test Based Accountability, and Educational Policy: A Scholarly Exchange Between Dr. Eric A. Hanushek and Drs. John Robert Warren and Eric Grodsky." *Educational Policy* 26:351–68.

Hanushek, Eric A., and Ludger Wössmann. 2006. "Does Educational Tracking Affect Performance and Inequality? Differences-in-Differences Evidence across Countries." *Economic Journal* 116(510):C63–C76.

_______. 2011 "The Economics of International Differences in Educational Achievement." In *Handbooks of the Economics of Education*, Vol. 3, edited by E. A. Hanushek, S. Machin, and L. Wössmann, 89–200. The Netherlands: North-Holland/Elsevier.

Horn, Daniel. 2009. "Age of Selection Counts: A Cross-Country Analysis of Educational Institutions." *Educational Research and Evaluation: An International Journal on Theory and Practice* 15(4):343–66.

Hout, Michael, and Stuart W. Elliott. 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 DC: National Academies Press.

Hoxby, Caroline. 2000. *Peer Effects in the Classroom: Learning from Gender and Race Variation*. NBER Working Paper 7867. Cambridge, MA: National Bureau of Economic Research.

Jones, M. Gail., Brett D. Jones, and Tracy Y. Hargrove. 2003. *The Unintended Consequences of High Stakes Testing*. Lanham, MD: Rowman and Littlefield Publishers.

Jürges, Hendrik, Kerstin Schneider, and Felix Büchel. 2005. "The Effect of Central Exit Examinations on Student Achievement: Quasi-Experimental Evidence from TIMSS." *Journal of the European Economic Association* 3(5):1134–55.

Kerckhoff, Alan C. 2001. "Education and Social Stratification Processes in Comparative Perspective." *Sociology of Education* 74(special issue):3–18.

LeTendre, Gerald K., Barbara K. Hofer, Hiteda Shimizu. 2003. "What Is Tracking? Cultural Expectations in the United States, Germany, and Japan." *American Educational Research Journal* 40(1):43–89.

Marks, Gary N. 2005. "Cross-National Differences and Accounting for Social Class Inequalities in Education." *International Sociology* 20(4):483–505.

_______. 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, Costas, and Mårten Palme. 2005. "Educational Reform, Ability, and Family Background." *American Economic Review* 95(1):414–24.

Oakes, Jeannie. 2005[1985]. *Keeping Track: How Schools Structure Inequality*. New Haven, CT: Yale University Press.

OECD. 2006. *Education at a Glance: OECD Indicators 2006*. Paris: Organisation for Economic Co-Operation and Development.

_______. 2007. *No More Failures: Ten Steps to Equity in Education*. Paris: Organisation for Economic Co-Operation and Development.

_______. 2009. *PISA 2006 Data Analysis Manual (SPSS)*. Paris: Organisation for Economic Co-Operation and Development.

_______. 2012. *PISA 2009 Technical Report*. Paris: Organisation for Economic Co-Operation and Development.

Pekkarinen, Tuomas, Roope Uusitalo, and Sari Kerr. 2009. "School Tracking and Development of Cognitive Skills." *IZA Discussion Paper No. 4058*.

Richer, Stephen. 1976. "Reference-Group Theory and Ability Grouping: A Convergence of Sociological Theory and Educational Research." *Sociology of Education* 49(1):65–71.

Rubin, Donald B. 1976. "Inference and Missing Data." *Biometrika* 63(3):581–92.

Scheerens, Jaap, and Roel Bosker. 1997. *The Foundations of Educational Effectiveness*. Kidlington/New York/Tokyo: Pergamon.

Schütz, Gabriela, Heinrich W. Ursprung, and Ludger Wössmann. 2008. "Education Policy and Equality of Opportunity." *Kyklos* 61(2):279–308.

Shavit, Yossi, Richard Arum, and Adam Gamoran. 2007. *Stratification in Higher Education: A Comparative Study*. Stanford, CA: Stanford University Press.

Shavit, Yossi, and Hans-Peter Blossfeld. 1993. *Persistent Inequality; Changing Educational Attainment in Thirteen 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: Clarendon Press.

Van de Werfhorst, Herman G., and Jonathan J. B. Mijs. 2010. "Achievement Inequality and the Institutional Structure of Educational Systems: A Comparative Perspective." *Annual Review of Sociology* 36:407–28.

Wilkinson, Ian A. G., Judy M. Parr, Irene Y. Y. Fung, John A. C. Hattie, and Michael A. R. Townsend. 2002. "Discussion: Modeling and Maximizing Peer Effects in School." *International Journal of Educational Research* 37(4):521–35.

Wössmann, Ludger. 2003. "Schooling Resources, Educational Institutions, and Student Performance: The International Evidence." *Oxford Bulletin of Economics and Statistics* 65(2):117–70.

Wössmann, Ludger, Elke Luedemann, Gabriela Schütz, and Martin R. West. 2009. *School Accountability, Autonomy, and Choice around the World*. Cheltenham, UK: Edward Elgar.

Wu, Margaret L., and Ray J. Adams. 2002. *Plausible Values: Why Are They Important?* Paper presented at the International Objective Measurement Workshop, New Orleans, LA.