

The high performance of Dutch and Flemish 15-year-old native pupils: explaining country differences in math scores between highly stratified educational systems

Citation for published version (APA):

Prokic-Breuer, T., & Dronkers, J. (2012). *The high performance of Dutch and Flemish 15-year-old native pupils: explaining country differences in math scores between highly stratified educational systems*. METEOR, Maastricht University School of Business and Economics. METEOR Research Memorandum No. 038 <https://doi.org/10.26481/umamet.2012038>

Document status and date:

Published: 01/01/2012

DOI:

[10.26481/umamet.2012038](https://doi.org/10.26481/umamet.2012038)

Document Version:

Publisher's PDF, also known as Version of record

Please check the document version of this publication:

- A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher's website.
- The final author version and the galley proof are versions of the publication after peer review.
- The final published version features the final layout of the paper including the volume, issue and page numbers.

[Link to publication](#)

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal.

If the publication is distributed under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license above, please follow below link for the End User Agreement:

www.umlib.nl/taverne-license

Take down policy

If you believe that this document breaches copyright please contact us at:

repository@maastrichtuniversity.nl

providing details and we will investigate your claim.

Download date: 26 Jan. 2022

Tijana Prokic-Breuer, Jaap Dronkers

The high performance of Dutch and Flemish 15-year-old native pupils: Explaining country differences in math scores between highly stratified educational systems

RM/12/038

METEOR

Maastricht University School of Business and Economics
Maastricht Research School of Economics
of Technology and Organization

P.O. Box 616
NL - 6200 MD Maastricht
The Netherlands

The high performance of Dutch and Flemish 15-year-old native pupils: Explaining country differences in math scores between highly stratified educational systems¹

Tijana Prokic-Breuer & Jaap Dronkers

Abstract

This paper aims to explain the high scores of 15-year-old native pupils in the Netherlands and Flanders by comparing them with the scores of pupils in countries with the same highly stratified educational system. Therefore, we compare only the educational performance of 15-year-old pupils from the following regions: the Netherlands, Flanders, Wallonia, the German Länder, the Swiss German cantons, and Austria. We use the data from the general Program for International Pupil Assessment (PISA) 2006 together with the specific PISA data of Germany and Switzerland also from 2006. We apply a multilevel model that takes into account the individual-, curriculum-, and system-level features in these highly stratified educational systems. The high scores of the Dutch pupils can be explained by the size of the Netherlands' vocational sector. The high Flemish scores can be only partly explained by the high curriculum mobility (as indicated by the lowest level of entrance selection). Central exit exams are not a good explanation of the high Dutch scores. Despite being limited to highly stratified systems, we still find educational policies and arrangements to have significant effects on the educational performance of pupils.

1. Introduction

Since 2000, 15-year-old pupils living in a large number of Organization for Economic Cooperation and Development (OECD) member states have been participating in the Program for International Pupil Assessment (PISA) every three years. The purpose of this assessment is to map competence in mathematics, physics, and reading at the end of the period of compulsory education (at the age of 15 or 16 years in most Western countries). Much attention was given to the high scores of the Finnish pupils in the three PISA waves, but the native pupils in the Netherlands and Flanders also scored quite high in reading and math relative to pupils in other European countries.

The Dutch and Flemish performance particularly stands out when compared to that of European countries with very similar educational systems and socioeconomic opportunities, such as Germany, Austria, and Switzerland. All these countries have highly stratified educational systems with a vocational education tradition (besides general education), which already influences the early stages of secondary education. In addition, the socioeconomic features of these countries are very similar; therefore the differences in the

¹ This research has been financed with the Professor Leune Award, which was bestowed in 2009 on the second author. We thank Cedgroep (www.cedgroep.nl), who made this award and prize possible. We also thank Mark Levels for his useful comments on an earlier version of this paper. All correspondence should be directed to the second author at j.dronkers@maastrichtuniversity.nl.

scores cannot be attributed to either system levels or compositional features. Consequently, our question is why do these pupils score differently in spite of these sociostructural and societal similarities?

This study aims to answer this and explain the high scores of 15-year-old pupils native² to the Netherlands and Flanders by comparing them with those of the pupils of countries with a highly stratified educational system, that is, Germany, Switzerland, and Austria.³ This restriction to highly stratified systems and native pupils is deliberate, for we are interested in the variation of outcomes and its determinants within the same type of system. Unlike previous studies that generally argue that different educational systems produce different outcomes (Shavit & Muller, 1998; Kerckhoff, 2001; Hanushek & Wössmann, 2005; Horn, 2009; Dunne, 2010), ours examines the differences *within* highly stratified educational systems in terms of both outcomes and determinants.

A particular feature of the Belgian, German, and Swiss educational systems is that their educational policies are fully decentralized. Each of the Belgian regions, German Länder, and Swiss cantons enjoys full political independence in their educational affairs. This autonomy is an ideal setup for studying the educational outcomes of pupils from the Netherlands and Flanders since it increases the number of highly stratified educational systems to which these two countries can be compared.

However, this approach has the hidden risk of producing results that are highly politically sensitive. Particularly in Germany, the comparison of the outcomes of the German Länder and their determinants is a delicate issue. One of the main reasons for such sensitivity is that Germany scores relatively low in cross-country comparisons, while the variation in average outcomes is very high across the German Länder. Consequently, the German Länder strictly forbids any internal German comparisons with the PISA and the consequences for disobeying this restriction are quite severe.⁴ To avoid such problems, we are forced to limit

² We limit ourselves in this study to native pupils because educational systems have different outcomes for pupils with a migrant background than for native pupils (Dronkers, van der Velden & Dunne, 2012) and because the Dutch PISA 2006 wave does not have information about the countries of birth of pupils and parents.

³ We did not include Liechtenstein or Luxemburg in our analysis, although these countries also have a highly stratified system, with general and vocational schools. Due to their small size and particular positions, inclusion would have posed problems.

⁴ The following is an excerpt of the contract we had to sign with the German educational authorities—represented by the *Institut zur Qualitätsentwicklung im Bildungswesen (IQB)*—to be granted access to so-called PISA+ data:

Data access is granted on condition that no German federal states are named or made identifiable – intentionally or unintentionally – in any publication of the research results. Furthermore, it is the responsibility of the data recipient to ensure that no German federal state can be identified through the combination of any parameters or features published with the research results.... For every case of non-compliance with his obligations set out in sentence 1, the data recipient shall pay the FDZ [*Forschungsdatenzentrum*] a contractual penalty of up to 10.000 €(ten thousand Euros). The right of the FDZ to claim further damages in the event of a higher amount of loss caused by the data recipient's offence remains unaffected.

our analysis and not explore existing differences between identifiable German Länder.⁵ Nevertheless, this is the first time this particular comparative setup has been used to explain the high scores of the Netherlands and Flanders.

Our approach to the question can be outlined as follows. We present the theoretical framework from which our hypotheses are derived. Next we outline the multilevel model with which we analyze the data. This model contains three levels (pupils, curricula, and systems) instead of the usual two levels approach (pupils and systems). Therefore we give this three-level model more attention. We then present the data and our analysis. The final section discusses the wider interpretation and consequences of our results, not only for the Netherlands and Flanders but also for countries with highly stratified educational systems (including Germany).

2. Differences in performance between educational systems: The trade-off between equity and efficiency

Only since recently have scholars agreed that the educational outcomes of pupils in different educational systems vary systematically. Pupils in comprehensive educational systems have, on average, higher scores than equivalent pupils in highly stratified educational systems, where the selection of students into different secondary education curricula takes place at a very early age (Shavit & Muller, 1998; Kerckhoff, 2001; Hanushek & Wössmann, 2005; Horn, 2009; Dunne, 2010).

The literature puts forward several arguments to explain this disparity. The degree of stratification of the educational system, measured by the degree of tracking, is directly related to a perceived trade-off between the equity and efficiency of educational achievement. Typically, highly stratified systems perform better in terms of efficiency and worse in terms of equity compared to comprehensive systems; however, the gain in efficiency is smaller than the loss in equity. In highly stratified education systems, where homogeneous classrooms permit a focused curriculum and appropriately paced instruction, maximizing learning within a track, the pupils in the lowest tracks will be systematically disadvantaged by slower learning environments that leave them behind those in the upper tracks in terms of skills.

⁵ This restriction on any internal German comparison of the educational performance of the Länder also applies to the NEPS data (National Educational Panel Study), the only educational data available at the German federal level (see <http://www.neps-data.de/en-us/home.aspx>).

Another explanation for the lower performance of pupils in highly stratified systems is the presence of so-called nonlinear peer effects. The diverse learning environment may benefit lower-ability students (in terms of higher motivation, exposure to better class discussion, etc.) without disadvantaging higher-ability students (Wössmann, 2003).

In sum, highly stratified systems have a high degree of efficiency from promoting more tailor-made learning; however, they produce lower overall equity by disadvantaging the students in the lowest tracks. The opposite is true for comprehensive systems.

In the production of lower or higher efficiency and equity schools play an important role as the mediating features between the systems and pupils. School characteristics, being a measure of educational environment, transmit educational system features into constraints and opportunities for pupils to learn and for teachers to teach. School features are the consequences of both the level of stratification of educational systems and the choices made in the design of any educational system, independently of its stratification level. For instance, differences in school socioeconomic composition exist in comprehensive educational systems (as a consequence of spatial segregation and school catchment areas) and these differences can be larger in highly stratified systems because of the selection of pupils with different scholastic abilities (related to their social background) into schools with different levels of curricula. However, if a more challenging curriculum level is taught in some schools within a highly stratified system, this can lead to the higher educational performance of the pupils in these schools. However, less able pupils will be selected into schools with less challenging curricula and their performance may be lower as well. Analogously, if only the average level of curriculum is taught in all schools within a comprehensive system, the more able pupils will be less challenged and thus may have a lower educational performance, while the less able students will be more challenged and thus have a higher performance. System and school features can thus form configurations that may be more or less of an advantage in the educational performance of pupils of different socioeconomic backgrounds attending schools with different compositions and offering different levels of curricula.

Dunne (2010) showed empirically that school characteristics have substantially different effects on and implications for educational achievement in different educational systems. These authors found that school composition matters: Pupils in schools of low socioeconomic composition in comprehensive systems have higher educational performance than those in comparable schools of low socioeconomic composition in highly stratified education systems, possibly because pupils within comprehensive systems are taught a more

challenging, average curriculum, while a less challenging, below-average curriculum is taught in schools within highly stratified systems. These differential effects of school composition in different educational systems may be explained by the selection of less or more able pupils into different school types (and thus different school compositions) in highly stratified systems, while such a selection of less or more able pupils into different schools does not exist in a comprehensive system (although these schools may still have a different composition due to spatial segregation).

In a follow-up study, Dronkers, van der Velden, and Dunne (2011) added curriculum level to their analysis of the effect of educational systems. The authors concluded that the inclusion of the curriculum level is necessary to avoid overestimation of the school composition effect. Especially in stratified educational systems such as those of the Netherlands and Germany, different tracks may be nested within the same school, such that the direct learning environment of pupils is better approximated by curriculum characteristics than school characteristics. But the addition of curriculum level did not alter the earlier conclusion that school characteristics have substantially different effects and implications in different educational systems regarding educational achievement.

Given that this study focuses only on a group of highly stratified educational systems, we start with the naïve assumption that individual socioeconomic outcomes or school features do not have significantly different educational returns across these countries because they all have the same educational system. If this is the case, then the next logical question is what other factors can explain these differences? This article's main argument is that there are important differences between highly stratified educational systems that can be used to explain the success of the Dutch and Flemish educational systems and thus the differences in achievement between the German Länder and the Swiss cantons. The next section outlines the main argument and put forward the hypothesis to be tested.

3. Systematic differences within highly stratified systems

There are three important features in which highly stratified educational systems can differ: 1) the *standardization* of the system by the existence of a central final exam, 2) the system's *vocational orientation*, and 3) the *selection* of pupils into different tracks at the start of secondary schooling and the level of their mobility between these tracks during secondary education. Interestingly, the current literature has devoted some attention to the relevance of these features, but they are often confused with the degree of stratification of the educational system (Shavit & Muller, 1998; Kerckhoff, 2001; Hanushek & Wössmann, 2005; Horn,

2009; Dunne, 2010). This omission risks overlooking the independent role these features have on the effects of the stratification of the educational system on individual achievement.

These three aspects varies between highly stratified systems and may consequently alter the trade-off differentiated educational systems face between equity and efficiency. Moreover, there is a reason to believe that these features can either 1) improve equity without reducing efficiency or 2) improve efficiency without decreasing equity in highly stratified educational systems. Essentially, the trade-off between equity and efficiency centers around the questions of which and how many pupils gain or lose in a particular educational system. The average outcome within each system reflects the net overall gain or loss in the educational performance of average pupils. Given that the average lower performance of the highly stratified system is a combination of the lower performance of the larger numbers of students in the lower tracks and the higher performance of the smaller numbers of students in the higher tracks, the way grouping of students in tracks occur within highly stratified systems becomes of critical importance in the trade-off between equity and efficiency.

This finding implies that the overall performance of students can be improved in two ways: 1) by decreasing selectivity and allowing more students to enter the highest track, which would increase the performance of middle track students because of the nonlinear peer effect, and 2) by increasing the educational focus of the lower tracks and attracting more able students to them. *Vocational orientation* increases the labor market focus of an educational system, particularly for students of the lowest tracks. Successful vocationalization of the lower tracks promotes efficient learning and employability and these tracks offer an attractive alternative to general education. Therefore, in terms of employability, we expect the quality and attractiveness of a vocational orientation to be directly related to the performance of pupils within the lowest tracks.

While the change in the relative sizes of the tracks increases with equity, it is *standardization* of the educational system that brings about gains in overall performance through higher efficiency in *all* tracks. The existence of a final central exam is believed to reduce variability between schools of the same type and level. By introducing an external standard, student performance is comparable across classes and schools and levels of educational quality are rendered observable and transparent. This improves the external monitoring of teachers, schools, and pupil performance. By setting such standards, a central examination promotes more equal opportunities and efficient learning through three different

channels: increasing external rewards for learning, decreasing peer pressure against learning, and increasing the monitoring of teachers and schools (Wössmann, 2003).

Our assumption is that the success of the Dutch and Flemish educational systems is best explained by successful implementation of the above-mentioned features. Table 2 shows that Netherlands and Flanders indeed stand out when compared to other countries regarding the degree of standardization, vocationalization, or mobility of their educational systems. Although the Netherlands score high on standardization and vocationalization, Flanders has the lowest level of entrance selection. The Netherlands has the highest score for the indicator central exit exam (average 0.3 in Table 1) and exceptionally prominent vocational education. Compared to other educational systems, that of the Netherlands has an exceptionally large vocational sector (53.3% compared to an average of 22.0%; see Table 1).

The Flemish educational system has a relatively open admission to each curriculum level of secondary school, but a high level of internal (downward) curriculum mobility (the “cascade model”) as well (Boone & Van Houtte, 2013). Table 2 shows that Flanders has a relatively low score on high selection entrance (0.15 compared to an average of 0.30; see Table 1) and a relatively high score on low entrance selection (0.31 compared to an average of 0.21; see Table 1). This may be a reflection of their unique cascade model, with a low level of entrance selection but a high level of internal downward mobility. Consequently, we formulate our two main hypotheses as follows.

1. The high scores of the Dutch pupils can be explained by a successful combination of vocationalization and standardization. We expect the Dutch educational system to be more successful at enhancing competition between schools and promoting efficient learning than other highly stratified educational systems because of the high level of parental school choice in the Netherlands (Netherlands has, together with Flanders, the highest percentage of pupils in private dependent schools, 69%, compared to an average of 42%). In the case of the Netherlands, competition is achieved by a combination of setting external standards for the quality of schooling (a final central test at the end of primary school and a final central examination for all types of secondary education) and maintaining an attractive and substantial vocational education. If this is true, then the Dutch educational system is more successful at reducing the lower performance of the less successful students (through vocationalization) than other highly stratified educational systems while increasing the efficiency of the system as a whole (through standardization).

2. Low selectivity (as indicated by low entrance selection) can explain the high scores of the Flemish pupils. We expect the Flemish educational system to promote equality among

its students without reducing its efficiency by placing a higher proportion of them in a higher track at the start of their secondary education at age 12. A unique feature of the Flemish system is that due to low selectivity, a majority of the students are allowed to enter the highest educational track. However, unlike for comprehensive systems, the existence of the lowest tracks ensures the possibility of downward mobility during secondary education. We argue that next to equity benefits related to the bigger size of the highest tracks, the motivation of students to stay in the highest track is higher than the motivation to exit from the lowest track; therefore the educational performance of all pupils can be increased.

4. Data, variables, and methodology

Since 2000, the OECD has conducted large-scale tri-annual tests among 15-year-olds living in its member states and partner states to assess pupils' mathematical, reading, and scientific literacy, resulting in the PISA data. The purpose of this test is to map competence in mathematics, physics, and reading at the end of the period of compulsory education (the age of 15 or 16 in most Western countries). We make use of the 2006 wave for the Netherlands, Belgium, German Länder, and the Swiss German cantons. The PISA 2006 data of the Swiss German cantons were purchased from the *Swiss Federal Statistical Office (EDK)*.⁶ The PISA 2006 data of the German Länder were obtained from the *Institut zur Qualitätsentwicklung im Bildungswesen (IQB)* of the *Humboldt University (Berlin)*. These additional and extended 2006 German PISA data were collected by the *Max-Planck-Institut für Bildungsforschung/IPN Kiel/the KMK*. The PISA data for Austria, Belgium, and the Netherlands were obtained from the PISA webpage. Belgium pupils are divided into Flemish and Walloon pupils based on the teaching language of their school.⁷

The PISA data for each participating country constitute a representative sample of the schools that teach 15-year-old pupils. We select only native-born pupils from these datasets, using the PISA definition of immigrant pupil.⁸ Each school that has been selected tests a sample of all 15-year-olds, irrespective of their curriculum or grade. In addition to educational performance, PISA also supplies information on a large number of individual background and school characteristics. The school principals provide details on a variety of school characteristics such as pupil–teacher ratio, teacher shortages, and school location. The

⁶ Swiss dataset of the IX-grade students, PISA 2006, – *Federal Statistical Office (EDK)*.

⁷ Schools in Brussels are thus divided into Flemish and Walloon schools.

⁸ The student and/or both parents were born outside the test country.

pupils' questionnaires request information on such things as the educational level of their parents. Table 1 gives the summary statistics of all the variables.

[About here table 1]

4.1. Combination of schools and curricula as an indicator of school level

The PISA data contain two cross-national indicators of the pupil curriculum. The pupil is asked whether he or she is currently enrolled in a certain curriculum of a certain level. This is later recoded in an internationally comparable format, distinguishing between general and vocational curricula, on the one hand, and between lower and higher curricula, on the other.

Schools are the sampling units in the PISA survey but can contain both general and vocational education. In addition, different definitions of a school are applied in different countries. Whereas in some, such as the Netherlands a school is defined as an “administrational unit”, in others, such as Germany, it is defined as a “location”. Therefore, we assume that the curriculum-level characteristics within these schools reflect more precisely the level of the daily teaching and learning environment than the administrative unit of a school. We call this combination of schools and curricula the curriculum level. We compute the indicator of the curriculum level for each pupil by combining his or her school identification number, the kind of curriculum he or she is following (vocational or general), and the curriculum level (low, medium, or high).

While focusing on highly stratified educational systems, this study also includes both school and curriculum features. Indeed, school and curriculum characteristics are not independent of the educational systems in which they operate but, rather, heavily conditioned by them. Therefore our analysis has three-levels: *educational systems* (either the unitary state or the cantons, Länder, or language community), *curricula* as a measure of learning environments and educational goals, and *pupils* with different social and cultural backgrounds and learning histories.⁹

4.2. Dependent variable: Math performance

The dependent variable in this study is *math performance*. To measure math skills accurately would make the test too long to be feasible. Hence a large number of very similar but shorter tests were created. Since such different tests can never have exactly the same degree of

⁹ We also added a fourth highest level for the five national states without any explanatory variable (see Section 5).

difficulty, *item response modeling* was used to achieve comparable results between pupils who took different tests. Our analysis averages the five plausible values obtained from the item response modeling. The math skills scores are standardized for the OECD countries, using an average of 500 and a standard deviation of 100. Table 2 shows the average scores and standard deviations of the reading and math scores by country, Belgium language community, Swiss canton, and combination of German Länder.¹⁰

[About here table 2]

4.3. Individual-level variables of all pupils

All the averages and standard deviations of these variables are given in Table 2 by country, Belgium language community, Swiss canton, and combination of German Länder.

4.3.1. Parental ESCS. The index of the economic, social, and cultural status (ESCS) of the parents is a composite index created in the PISA dataset of the occupational status of the parents measured with the International Socio-economic Index for Occupational Status (ISEI) scale (Ganzeboom, De Graaf, Treiman & De Leeuw, 1992), the educational level of the parents measured with the International Standard Classification of Education (ISCED; UNESCO, 2006), and the presence of any material or cultural resources at the pupils' homes.¹¹ This combination of the parents' occupational status and educational level together with the resources at home produces the strongest indicator of the parental environment. We set the average of the parental ESCS for each destination country to zero to ensure that the comparisons for this item show the result for the average pupil in these countries.

4.3.2. Grade. We have included a variable to account for the fact that not all pupils attend the same grade. As a result of between-country variance in the way grades are constructed, we have standardized the grade around a country's modal grade.

4.3.3. Female. We control for gender effects by using a dummy variable indicating whether a pupil is female (one) or male (zero).

¹⁰ Comparable detailed information about the averages of the Länder has already been published by Köller, Knigge, and Tesch (2010). However, according to the contract (see footnote 3), we are not allowed to publish comparable detailed information. Therefore we grouped the Länder into five categories, but only for this table, not in the analysis.

¹¹ The measure consists of the presence of a desk, a private room, a quiet place to study, a computer, educational software, Internet access, literature or poetry, art, books that may be of use in schoolwork, a dictionary, a dishwasher, and the presence of more than 100 books in the house.

4.4. Variables measured at the curriculum level

All the averages and standard deviations of these variables are given in Table 2 by country, Belgium language community, Swiss canton, and combination of German Länder.

4.4.1. *Levels of curriculum.* We distinguish between three curriculum levels: 1) *Lower vocational curriculum*, a dummy indicating whether a pupil is currently enrolled in lower vocational type of education (ISCED classification); 2) *Medium*, a dummy indicating whether a pupil is currently enrolled in a higher vocational or lower general type of education, and 3) *High*, a dummy indicating whether a pupil is enrolled in the highest level of general education, giving entrance rights to university education. In addition, we use these dummies as our indicators of curriculum levels to compute school composition features. More information on these curriculum levels are in Appendix I.¹²

4.4.2. *Socioeconomic school diversity.* Similarly, we calculate the sociocultural diversity at the curriculum level. Using the ESCS scores of the parents, we divide these parental scores into five categories: the group with the lowest 10% of scores, the 10–30% group, the 30–70% group, the 70–90% group, and the group with the highest 10% of scores.¹³ On the basis of these five categories, we calculate the Herfindahl index of sociocultural diversity (varying between zero and one).¹⁴ The index should be interpreted as follows: A value of zero means there is no diversity because all parents of all pupils at that particular curriculum level are in the same ESCS category. A value approaching one indicates a very high level of diversity, where pupils are equally recruited from the five ESCS categories. Since this Herfindahl index of sociocultural diversity is “level blind” and therefore insensitive to the average parental educational level, we added the average ESCS of the curriculum level to the analysis.

4.4.3. *Percentage of native pupils.* The Dutch PISA data 2006 do not contain the country of birth of the pupils and parents because the Dutch Minister of Education did not want to know that important variable (Dronkers, 2005). Therefore we cannot distinguish between immigrants from different origin countries. In this case the omission is less serious than in an OECD-wide comparison, because the countries of origin of the immigrants in Austria, Belgium, Germany, and the Netherlands are more or less comparable (see Appendix C of

¹² Note that the construction of these three dummies deviates from the approach of Dronkers, van der Velden & Dunne (2011), who divide the Medium category into a vocational and a general type.

¹³ The groups are defined as follows: 1) less than 10%, $ESCS < -1.1$; 2) 10–30%, $-1.0 < ESCS < -0.4$; 3) 30–70%, $-0.3 < ESCS < 0.6$; 4) 70–90%, $0.7 < ESCS < 1.2$; 5) more than 90%, $ESCS > 1.3$.

¹⁴ The Herfindahl index of sociocultural diversity is calculated as follows: $1 - ((\text{percentage of parents from ESCS group 1})^2 + (\text{percentage of parents from ESCS group 2})^2 + \dots + (\text{percentage of parents from ESCS group 5})^2)$.

Dronkers, van der Velden & Dunne, 2011). However, the percentage of native pupils in each of these educational systems does vary and needs to be included in our analysis.

4.4.4. Average ESCS. We calculate the average parental ESCS per curriculum level as the best indicator of school socioeconomic composition.

4.5. Variables measured at the school level

All the averages and standard deviations of these variables are given in Table 2 by country, Belgium language community, Swiss canton, and combination of German Länder. We use these school-level indicators as curriculum-level indicators.

4.5.1. Selective admittance. Selective admittance of pupils to a school is a scale in the PISA data based on the answers of the principals indicating whether admittance to their schools is based on academic record and/or recommendations. We divide the scale into three dummies: high selection, medium selection (reference category), and low selection.

4.5.2. Urbanization. Two dummies were constructed to indicate whether a school is located in (large) *city* or a *rural area*. Schools in an urbanized countryside or a (small) town serve as the reference category.

4.5.3. School size. School size is the number of pupils in the school, based on the answers of the principals.

4.5.4. Private/public. Educational systems differ in the shares of public and private schools and the degree of state grants for these private schools. Two dummies are constructed to separate private state dependent and private state independent schools from public schools. These variables control for these system differences and the effectiveness of these school types (Dronkers & Robert, 2008; Dronkers & Avram, 2010a, 2010b).

4.6. Variables measured at the educational system level

All the averages and standard deviations of these variables are given in Table 2 by country, Belgium language community, Swiss canton, and combination of German Länder.

4.6.1. Central exit examination. This standardization variable refers to the degree to which clear external standards or incentives exist in an educational system to maintain the quality level of the education it provides. This variable indicates the extent to which educational systems have a set of standard rules and guidelines education should follow (Wössmann, 2003). An example of a clear external standard is conducting nationally standardized exams at the end of secondary education. Therefore, we have constructed the dummy variable *central exit examination*, which takes value one if there is a standardized

central exit exam at the end of high school in the educational system. This dummy takes the value zero if there is none.

4.6.2. Percentage of pupils enrolled in a lower vocational curriculum. Studies (Shavit & Blossfeld, 1995) have shown that an *educational system's vocational orientation* increases its labor market focus. This can reduce (vocational education as a dead-end street) or enhance (vocational education as a safety net) equality and promote efficient learning. We control for this vocationalization of each educational system by looking at the percentage of the total number of pupils enrolled in the lower vocational curriculum. We derive this percentage by aggregating the percentage of pupils in lower vocational education by country, region, Land, or canton.

4.6.3. Socioeconomic gradient. This variable depicts the relation between the educational achievement and socioeconomic background of the pupils in each separate educational system. The gradient by country, region, Land, or canton is derived from the coefficient of the ESCS after controlling for the individual characteristics of gender and grade (Appendix II, model 2, but separately by country, region, Land, or canton). A high score of this socioeconomic gradient means that the effect of parental background on educational performance is strong in that country, region, Land, or canton.¹⁵

4.6.4. First age of selection of pupils. This age varies only between 10 and 12 years within highly stratified systems.

4.7. The Model

We use multilevel modeling for the purpose of our empirical analysis because our data are hierarchically clustered in groups. Therefore the data violate the conditional independence assumption about error terms from the classical linear regression model. This multilevel analysis has four hierarchical levels: the national state, educational systems, curriculum levels, and the pupils. In a number of cases the multilevel solution did not converge. We indicate such a non-convergence in Tables 4 and 5. In these cases we applied a three-level model (educational systems, curriculum levels, and pupils).¹⁶

¹⁵ Controlling for grade can lead to an underestimation of the socioeconomic gradient of a country, region, Land, or canton if repeating or skipping classes in a country is more dependent on socioeconomic background than in the other countries. Given the similarities of the highly stratified systems analyzed, substantial cross-country variation in the relation between repeating or skipping classes and socioeconomic background is unlikely.

¹⁶ The non-convergence of some four-level models may be due to the limitation of the general program STATA, which we had to use to analyze the German data in the IQB building. The MIWin program, specialized in multilevel analysis, converges with no problem in situations where STATA does not converge but we were not allowed to use our laptops with MIWin for this analysis in the IQB building.

We divide the observations into five different groups at the national state level according to the national state to which a student belongs: Austria, Belgium, Germany, the Netherlands, or Switzerland. We have added this national state level to capture the possible effect of a specific national state on educational performance, besides the features of the different educational systems within these states. Note that in all models this national state level contains significant unexplained variance. Only for Austria and the Netherlands is there no variation at the educational system level, because both countries are unitary states with regard to education. The dummy for the Netherlands is thus situated at the national state level, with only five cases. The dummy for Flanders is situated at the educational system level, with 26 cases. School characteristics are transferred to the curriculum level to make the multilevel analysis manageable. The advantage of multilevel modeling for the purpose of our analysis is that it allows for a direct estimation of the importance of these different levels in the causal process.

5. Analyses

Table 2 shows the summary statistics for all relevant individual, school, and system variables for the different educational systems of the countries, regions, cantons and groups of Länder, separately. Table 3 shows the differences in the average level of ESCS for each level of curriculum.¹⁷ As we can see from these tables, although these educational systems are all highly stratified, with a strong tradition in vocational training, there are considerable differences between them regarding the average ESCS at the individual and curriculum levels. Moreover, the systems differ in the socioeconomic segregation of pupils at different levels.

[Table 3, about here]

We conducted a series of multilevel analyses for math scores. The full results are given in the Appendices II to V. Model 1 includes only dummies for the Netherlands and Flanders and thus gives the degree to which the scores of Dutch and Flemish pupils are higher than the average scores of pupils from Austria, the German Länder, the Swiss German cantons,

¹⁷ Because of the restrictions in the contract with IQB, we are not allowed to identify the Länder in Table 3. Therefore we averaged the values for all the German Länder into five groups. This grouping of Länder is not used in the multilevel analyses, where we use the Länder separately.

and Wallonia.¹⁸ Model 2 includes the Dutch and Flemish dummies and the individual characteristics.

[About here table 4]

Models 3 to 6 include the curriculum-level features separately, while model 7 contains controls for both individual and all curriculum features together. We summarize the coefficients for Flanders and the Netherlands from the multilevel equations in Table 4 and Appendix II gives all the coefficients of these seven models. Table 5 shows all the coefficients of model 7.

The next step in our series of multilevel analyses includes features of the educational systems. At this level we control for the existence of central exams, the size of vocational education, age of first selection, and the socioeconomic gradient. Each of these educational system features is combined with only individual-level features of model 2 in models 8 to 11. The coefficients of these four models are given in Appendix III. The results in Appendix IV (models 12 to 15) shows the same equations but with entrance selection as an additional control. The effects for the dummies for the Netherlands and Flanders in these models are presented in Table 4.

[About here table 5]

With models 16 to 23, we combine educational system features in the equations together with individual characteristics and different combinations of curriculum-level features. We present these various models and combinations to make our results as transparent as possible. All the coefficients of these models are given in Appendix V.

¹⁸ However, there is some uncertainty about the interpretation of the results obtained for the dummy for Netherlands. Given that multilevel modeling, just as normal ordinary least squares regression, requires a minimum number of observations at each level, we can question if this condition is met in the case of the Netherlands. At the highest level of our model (the nation state), where the Netherlands is, we have only five cases in which 26 educational systems are clustered. At the level of educational system, where Flanders is, we have 26 cases in which many schools are clustered. Because of the very low number of groups at the highest level, the meaning of the insignificance of any variable at this level can be questioned. Statistically, it may be that such insignificance occurs because of the low number of cases at that level and not because the coefficient is not substantive. On the other hand, we can be certain about significant effects found at this level because of the low likelihood of obtaining significant effects with fewer groups at each given level. Therefore, in the explanation of our results in the following sections, we will also take into consideration the size of the effect for the Dutch dummy, even if, strictly speaking, this should not be considered. However, the Dutch dummy is neither significant in a three level analysis.

6. Results

6.1. Explanations of the high scores of the Dutch and Flemish pupils

The main aim of this study is to determine why the pupils in the Netherlands and Flanders perform better in international educational comparisons than those in countries with very similar socioeconomic as well as educational system features. Our main argument is that in the Dutch case this success is achieved through vocationalization and standardization, and in the Flemish case through the insurance of low selection into the higher tracks.

We find mixed support for our first hypothesis on the success of the Dutch educational system, that is, the high scores of the Dutch pupils can be explained by a successful combination of vocationalization and standardization. When it comes to the role of the standardization of the Dutch educational system, we observe that the strength of the Netherlands dummy is only a bit lower¹⁹ in model 8 (including individual and exam variables) than in model 2 (including only individual variables). However, the Netherlands dummy and the effect of the central exit exam variable are both insignificant in the final model 24. This means that the existence of a central examination is not a good empirical explanation of the high scores of the Dutch pupils.

We do find that *the size of the vocational sector can explain the high scores of the Dutch pupils*. The strength of the Netherlands dummy is only a bit lower²⁰ in model 11 (individual variables and percentage vocational education) than in model 2 (only individual variables). But the effect of the percentage vocational education is positive and significant in the final model 24 (a 10% larger vocational sector give 11.4 points higher on the math test), while the Netherlands dummy is insignificant and small (15 points) and thus the percentage vocational education can explain the high Dutch math scores. The interim results from Appendices IV (model 15) and V (model 23) suggest that this positive effect of percentage vocational education becomes only significant after the inclusion of other controls (e.g., entrance selection, the curriculum level).

Moreover, Table 3 shows a related aspect of the size of the vocational sector: The average socioeconomic composition of Dutch vocational schools is relatively high compared to other educational systems; only the vocational schools in the Swiss canton Aargau have an even better socioeconomic composition. That might explain why in model 2, after the inclusion of the parental ESCS, the Netherlands dummy already becomes insignificant. Because the Dutch vocational sector both is among the largest of the compared systems and

¹⁹ A value of 29 points versus 39 points on the math test.

²⁰ A value of 30 points versus 39 points on the math test.

has the best socioeconomic composition of the lowest vocational sector, the Netherlands dummy already becomes small in model 5 (where we do not control for percentage vocational education or attending a vocational curriculum but already control for the socioeconomic background of the pupils).

Our second hypothesis, which assumes that *the low entrance selection can explain the high scores of the Flemish pupils*, is supported by our results. The Flemish dummy in model 4, with the inclusion of the variable entrance selection, becomes insignificant and smaller.²¹ The Flanders dummy is also insignificant in models 12 to 15 (Appendix IV), which have entrance selection as a control. As said earlier, Flanders has a relatively low score on high entrance selection and a relatively high score on low entrance selection. This may reflect their unique cascade model in secondary schools, with a low level of entrance selection but a high level of internal downward mobility (Boone & Van Houtte, 2013).

6.2. The difference that vocational orientation, standardization, and low entrance selection make

A low level of entrance selection still has a negative effect on educational performance (-3.6), while a high level of entrance selection has a positive effect (7.8). It may be that, despite all the controls, these effects still reflect some unmeasured variance between the pupils who have passed these various levels of entrance selection. But a more plausible explanation is that a school's level of entrance selection indicates its ambition and quality, which influence the performance of its pupils.

We do not find a positive effect on educational performance for a highly stratified system with central exit exams. As long as we do not control for the curriculum level, a central exit exam has a positive effect on math scores of around 12 points (model 8 or 16). It may be that within highly stratified educational systems, curricula and exit exams have the same function: setting standards for teachers and pupils. A curriculum may fulfill this function better and make the additional effect of a central exit exam insignificant.

Age of selection in a highly stratified system still has a positive effect on educational performance, despite all the controls. Given that the range of this age among these highly stratified educational systems runs between 10 and 12 and the German Länder apply in most cases 10 years, the two-year difference with Flanders, the Netherlands, and most Swiss German cantons gives German and Austrian pupils a 22.2 lower average math score.

²¹ A value of 25 points (model 4) versus 39 points (model 2).

The size of the vocational sector of a highly stratified system has a positive effect on the math scores of all pupils. Given the negative effect of attending a lower vocational track (-56.8), the arrear of these 57 points is bridged if the size of the vocational sector is equal to 50%.²² However, the scores of pupils in the medium and high curricula increase at the same pace with a larger vocational sector and thus these scores also increase by 57 points. This can be explained by the higher selectivity of the schools offering medium and high curricula if they receive lower percentages of less-able pupils due to the larger size of the vocational sector.

6.3. Other results

It should be noted that these results are also relevant for the individual German Länder because they are separately included in the multilevel analyses and not as the combinations presented in Tables 2 and 3. We expect that these results would not be different if we could openly compare these Länder, as we have done with Flanders and the Netherlands.

The parental socioeconomic background has its usual significant effect (7.5), but the strength of the average ESCS school composition (which uses the same index and has more or less the same range; see Table 1) in the final model 24 is three times higher (26.3). Attending a school with a high socioeconomic composition remains important for educational performance, despite the fact that we control for the curriculum level, a parameter as substantial as school composition (lower vocational track -56.8, high 55.0). Taking into account school curriculum level is thus also important in understanding the educational performance of pupils, but the attended curriculum level does not fully explain the effect of parental ESCS or the average ESCS school composition.

In view of the many policy debates about the benefits of ESCS school diversity for educational performance (as part of the debate on the need to change toward a more comprehensive educational system), it is remarkable that there is no significant effect of ESCS school diversity after controlling for curriculum level. It may be possible that ESCS school diversity is erroneously mixed up with the curriculum level and that the level effect is wrongly attributed to school diversity. It should be noted, however, that ethnic school diversity has a negative effect on the educational performance of migrant pupils, even after all possible controls (Dronkers & van der Velden, 2012). But this negative effect is contrary to the assumed positive effect of diversity.

²² Here 56.84 (attending a lower vocational stream)/1.139 (size of vocational sector).

The percentage of native pupils in schools has a positive effect on the educational performance of native pupils (10% more native pupils is related with 1.5 points more at the math test), despite all controls for ESCS, school composition, curriculum level, urban living, private schools, and so forth. A possible explanation is that a higher percentage of non-native pupils is a larger challenge for the teacher and thus decreases the real teaching and learning time for native pupils as well.

Pupils in private dependent schools still have, despite all controls, higher scores (by 3.2 points), while pupils in private independent schools have lower scores (by -18.5 points) than comparable pupils in comparable public schools. This is in line with earlier research (Dronkers & Robert, 2008; Dronkers & Avram, 2010a, 2010b), which found the same effects for these countries with highly stratified educational systems.

A large school size promotes the educational performance of its pupils (100 more pupils means a 2.5-point higher score). This result runs against lay wisdom that small schools are better for children but supports the general line of research. However, the negative effect of the school size squared variable indicates that this positive effect levels off and becomes negative for secondary schools larger than 4016 pupils.

Finally, the variance in a model with only the Netherlands and Flanders dummies (model 1, Table 5) can be divided between four hierarchical levels: the national state, educational systems, curricula, and pupils. The variances at these levels are as follows in model 1: 447 (5.5%), 87 (1.1%), 4132 (51.2%), and 3407 (42.2%). Our multilevel models can explain substantial parts of these variances, but the left variances of model 24 show that significant variances remain unexplained at each level—126 (3.3%), 130 (3.5%), 689 (18.4%), and 2809 (74.8%). This means that the observed educational system, school, and pupil characteristics decrease the unexplained variance at the national state level by 72% ($1 - 126/447$), at the school level by 83% ($1 - 689/4132$), and at the pupil level by 18% ($1 - 2809/3407$). However, these characteristics increase the unexplained variance at the educational system level by 33% ($87/130$). This means that these differences in these characteristics between the analyzed systems decrease the differences in the added value of the educational systems of the different countries, regions, cantons, and Länder.

7. Conclusions

The main aim of this article is to explain the high scores of 15-year-old pupils in the Netherlands and Flanders by comparing them with those of pupils in countries with similar highly stratified educational systems. Moreover, we examine if individual-, curriculum-, and

educational system-level features play different roles for educational achievement in highly stratified educational systems with general and vocational training. To this end, we compare the educational performance of 15-year-old pupils from the Netherlands, Flanders, Wallonia, the German Länder, the Swiss German cantons, and Austria.

We used multilevel modeling for the purpose of our empirical analysis because our data are hierarchically clustered in groups. Our multilevel analysis has four hierarchical levels: the national state, educational systems, curricula, and pupils. Different from previous studies about the effects of educational systems, we use curriculum level as the school level. We do so because it allows us to include the level of the curriculum taught in the analyses. We thus avoid overestimating the school composition effect, especially in stratified educational systems, or misspecifying the effects of the features of educational systems.

7.1. Explanation of the high Flemish and Dutch scores

The high scores of the Dutch pupils can be explained by the size of the Netherlands' vocational sector (first hypothesis). We find that the larger the vocational sector of a highly stratified system, the better all pupils within that educational system perform. The Netherlands has the largest vocational sector of all educational systems compared in this analysis. A large vocational sector may increase the labor market focus of an educational system and thus promote efficient learning by pupils through a more focused curriculum and teaching. Consequently, after the inclusion of this size of the vocational sector in the analysis, the scores of the Dutch pupils were the same as those for the other countries with highly stratified systems. The policy drive toward the generalization of the content of secondary education since 1968 (in an attempt to lower educational inequality) has threatened to make vocational education a dead-end street with a very low socioeconomic composition. Relative to policy makers in other highly stratified systems (such as Aargau, the German city-Länder, the northwestern and eastern German Länder, and Wallonia), the Dutch policy makers were unsuccessful in this policy drive and we contribute the high scores of all Dutch pupils to the large Dutch vocational sector.

Central exit exams are not an independent explanation of the high scores of Dutch pupils (first hypothesis). The curriculum may set the standards for teachers and pupils better within a highly stratified system than a central exam can.

The high Flemish scores can be partly explained by the high curriculum mobility (as indicated by the highest level of medium entrance selection). The Flemish educational system has relatively open entrance at each curriculum level in secondary school, but a high

level of internal (downward) curriculum mobility (“cascade model”) as well. The “not too high but not too low” level of entrance selection (trying to combine the best of two solutions) and the high level of curriculum mobility within schools and between tracks improve the matching of pupils to their educational attainment and achievement. This can improve efficient learning and thus leads to high scores. However, we cannot fully explain the high scores of the Flemish pupils. A historical explanation may be that the successful 20th-century struggle for survival of the Flemish language may have given the Flemish education a higher ethos and desire to outperform their Walloon co-patriots.

7.2. Educational policy can make a difference

Our analysis shows another very important result, with further implications than only for Dutch and Flemish policy makers. Despite our limitation to only countries, regions, Länder, or cantons with highly stratified systems, we still find significant effects of educational policies and arrangements on the educational performance of pupils. This means that policy makers do not need to first overhaul their educational system from a highly stratified one into a comprehensive one before they can improve the educational performance of their country. It also means that parents and teachers can more easily build coalitions between citizen and schools to improve their schools and that they do not need to wait until the educational system can be changed.

We briefly summarize which policies and arrangements within highly stratified system can make a difference. Some *entrance selection* by schools can be useful to strengthen their ambition and quality, which influence the performance of their pupils. The introduction of *private dependent schools* (in contrast with private independent schools) can increase a moderate form of competition between schools for pupils and thus improve teaching quality. The small *size of secondary schools* can be negatively related to educational performance, but building large “education factories” can also be harmful. *Curricula and exit exams* are important for the same function: setting standards for teachers and pupils regarding what to teach. Curricula seem to fulfill this function better than exams. *First selection* at the age of 12 instead of 10 increases the educational performance of all pupils. A *vocational sector* that is large enough to be vibrant and promote efficient teaching in preparation for the labor market improves the educational performance of all pupils, including those who do not attend vocational education. The related higher selectivity of schools offering medium and high curricula if they receive lower percentages of pupils due to the larger size of the vocational sector may avoid a watering down of the curriculum and a

slackening of standards. We do not find a trade-off between *socioeconomic inequality* and high educational performance within highly stratified systems, in contrast to Von Below (2002).

7.3. Implications for the explanation of the differences in educational performance of the German Länder

Although the intention of this paper is not to compare the educational systems of the German Länder, some conclusions about the effectiveness of their educational systems can be drawn from the analysis. It is the first sophisticated analysis of the effectiveness of highly stratified systems for the educational performance of native pupils.²³ Of the 26 educational systems compared, 14 are part of the *Bundes Republik Deutschland* and the educational systems of the other non-German countries have many features in common with it.

Some *entrance selection* by schools can be useful to strengthen their ambition and quality, which influence the performance of their pupils. Compared with low, medium, and high entrance selection average values of 21%, 49%, and 30%, respectively, the city-Länder have a remarkable low level of high entrance selection (9%) and the same holds for the northwestern Länder (15%). Given the positive effect of high entrance selection on math scores, these low levels of entrance selection may be detrimental for the educational performance of native pupils in these Länder.

Private dependent schools (in contrast with private independent schools) can increase a moderate form of competition between schools for pupils and thus improve teaching quality. The average percentage of private dependent schools is 23%, and the Netherlands and Flanders are mainly responsible for this average. All German Länder have lower percentages, especially the northwestern and northeastern Länder.

First selection at the age of 12 instead of 10 increases the educational performance of all pupils. All German Länder have 10 years as the age of first selection, which is quite early compared with the other countries with highly stratified systems. Increasing the age of selection by one year may increase the math scores of German native pupils by 11 points.

A *vocational sector* that is large enough to be vibrant and promote efficient teaching in preparation for the labor market improves the educational performance of all pupils, including those who do not attend vocational education. The related higher selectivity of

²³ The only available comparison between the German Länder (Köller, Knigge & Tesch, 2010) applies descriptive tables and simple OLS regressions, does not include educational system characteristics to explain cross-Länder differences and does not distinguish between native pupils and pupils with a migrant background.

schools offering medium and high curricula if they receive lower percentages of pupils due to the larger vocational sector may avoid watering down the curriculum and slackening standards. Although vocational education is often seen as a German institution, the size of the vocational sector in northeastern Länder (5%) is below the average (13%).

Higher percentages of native pupils in schools increase the math scores of native pupils (10% more native pupils, 1.5 points higher). The city-Länder have a lower percentage of native pupils (74%) than other Länder, such as the northeastern Länder (96%) and southeastern Länder (91%), with an average percentage of 87% native pupils for all highly stratified systems compared. This variation in the percentage of native pupils means that the lower math score of the city-Länder (508) is an underestimation of the real quality of their educational organization, while the high math score of the southeastern Länder (535) overestimates the real quality of these Länder' educational organization.

One of the reasons for the ongoing stalemate in the German debates of educational policies is the assumed trade-off between *socioeconomic inequality* and high educational performance within highly stratified systems (Von Below (2002)). We do not find such a trade-off for the educational systems compared. The same holds for the German Länder, whose average socioeconomic gradient is 9.4; the gradient for the five combinations of German Länder is 8.1 for the northeastern Länder, 8.9 for the southeastern Länder, 11.1 for the city- and southwestern Länder, and 12.1 for the northwestern Länder.

Unfortunately, discussing the policy implications for the German Länder in more detail is forbidden. But already the results show that a more transparent analysis of the differences in the educational performance of the Länder will enlighten the German public and can break the ongoing stalemate in German educational policy making through independent research. The remaining unexplained variance in the educational system levels shows that the added value of the education of the German Länder will be quite different and thus interesting for parents, teachers, and policy makers interested in a transparent state.

References

Boone, S. & Van Houtte, M. (2013) Educational choice at the transition from primary to secondary education: Further evidence for a rational calculation? *Sociological Review* (in press)

Dronkers, J. (2005). Opnieuw een gemiste kans met PISA [Again a problem with the Dutch PISA data]. *Het Onderwijsblad*, April 2.

Dronkers, J. (2010). Features of educational systems as factors in the creation of unequal educational outcomes, pp. 299–328, in *Quality and Inequality of Education. Cross-National Perspectives*, edited by J. Dronkers. Dordrecht/Heidelberg/London/New York: Springer.

Dronkers J. & Avram, S. (2010a). A cross-national analysis of the relations of school choice and effectiveness differences between private-dependent and public schools. *Educational Research and Evaluation* 16: 151–175.

Dronkers, J. & Avram, S. (2010b). A cross-national analysis of the relations of school choice and effectiveness differences between private-independent and public schools. *Sociological Theory and Methods* 25: 183–206.

Dronkers, J. & Roberts, P. (2008). Differences in scholastic achievement of public, private government-dependent, and private independent schools: A cross-national analysis. *Educational Policy* 22: 541–577.

Dronkers, J. & van der Velden, R. (2012). Positive but also negative effects of ethnic diversity in schools on educational performance? An empirical test using cross-national PISA data, edited by M. Windzio, in *Integration and Inequality in Educational Institutions*. Dordrecht /Heidelberg/ London/New York: Springer.

Dronkers, J., van der Velden, R. & Dunne, A. (2011). The effects of educational systems, school-composition, curriculum-level, parental background and immigrants' origins on the achievement of 15-year-old native and immigrant pupils. A reanalysis of PISA 2006. Maastricht University, ROA Research Memorandum 2011/6.

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: 11-44.

Dunne, A. (2010). Dividing lines: Examining the relative importance of between- and within-school differentiation during lower secondary education. Ph.D., European University Institute, Florence.

Hanushek, E.A. & Wössmann, L. (2005). Does educational tracking affect performance and inequality? Differences-in-differences evidence across countries. *Economic Journal* 116: 63-76.

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

Kerckhoff, A.C. (2001). Education and social stratification processes in comparative perspective. *Sociology of Education*, 74 (Extra Issue): 3–18.

Köller, O., Knigge, M., & Tesch, B. (Eds.) (2010). *Sprachliche Kompetenzen im Ländervergleich*. Münster: Waxmann.

Scheerens, J. & Bosker, R. (1997). *The Foundations of Educational Effectiveness*. Kidlington/New York/Tokyo: Pergamon.

Shavit, Y. & Blossfeld, H.-P. (Eds.) (1993). *Persistent Inequalities: a Comparative Study of Educational Attainment in Thirteen Countries*. Boulder Colorado: Westview Press.

Shavit, Y. & Müller, W. (Eds.). (1998). *From School to Work. A Comparative Study of Educational Qualifications and Occupational Destinations*. Oxford: Clarendon Press.

Von Below, S. (2002). *Bildungssysteme und soziale Ungleichheit*. Berlin: Lehmanns.

Wößmann, L. (2003). Central exit exams and student achievement: International evidence. In M. West & P. Peterson (Eds.), *No Child Left Behind? The politics and practice of school accountability* (pp. 292-323). Washington, DC: Brookings Institution Press

Table 1: Summary statistics of all educational systems pooled together

Variable	No. of Observations	Mean	Std. Dev.	Min	Max
Math score	48,786	529.73	86.93	150.09	1072.80
Parental ESCS	48,786	0.35	0.81	-3.68	3.20
Female	48,786	0.49	0.50	0	1
Grade	48,385	0.04	0.60	-2.74	2.87
<i>Curriculum features</i>					
Average school ESCS	48,786	0.30	0.44	-2.25	1.69
ESCS school diversity	48,786	0.64	0.08	0	0.78
% native pupils curriculum	48,786	86.88	14.30	4.55	100
Lower vocational curriculum	48,786	0.22	0.41	0	1
High curriculum	48,786	0.34	0.47	0	1
Low selection	48,786	0.21	0.41	0	1
High selection	48,786	0.30	0.46	0	1
Private gov. independent	48,786	0.01	0.10	0	1
Private gov. dependent	48,786	0.23	0.42	0	1
School size	48,786	655.99	397.07	15	3931
Rural area	48,786	0.39	0.49	0	1
City	48,786	0.21	0.41	0	1
<i>Educational system features</i>					
% lower vocational education	48,786	22.03	13.18	0	53.31
Age of selection	48,786	10.79	0.96	10	12
Exit exam	48,786	0.32	0.47	0	1
Socioeconomic gradient	48,786	9.44	4.58	0.57	25.05

Source: PISA 2006. The data for the Swiss cantons are from the Swiss Federal Statistical Office (EDK), the data for the German Länder are from the *Institut zur Qualitätsentwicklung im Bildungswesen* (IQB), and the data for all the other countries are from the PISA website.

Table 2: Summary statistics of all variables for each of the educational systems

	Math Score	Parental ECSC	Female	Grade	% Native Pupils	Average School ESCS	ESCS School Diversity	Lower Vocational Curriculum	High Curriculum
Aargau	557.80	0.25	0.48	0.03	82.62	0.15	0.65	0.15	0.45
	(79.84)	(0.78)	(0.50)	(0.66)	(14.72)	(0.42)	(0.09)	(0.36)	(0.50)
Bern (German)	544.23	0.13	0.51	-0.13	90.28	0.09	0.62	0.34	0.24
	(75.58)	(0.82)	(0.50)	(0.57)	(12.77)	(0.46)	(0.08)	-(0.48)	-(0.43)
Basel-Landschaft	541.88	0.27	0.50	-0.02	84.58	0.21	0.63	0.27	0.34
	(81.64)	(0.77)	(0.50)	(0.61)	(13.10)	(0.40)	(0.09)	(0.44)	(0.47)
St. Gallen	564.38	0.09	0.49	0.00	81.13	0.02	0.66	0.30	0.15
	(75.93)	(0.83)	(0.50)	(0.66)	(13.54)	(0.39)	(0.08)	(0.46)	(0.36)
Schaffhausen	577.70	0.29	0.49	0.01	82.54	0.22	0.64	0.31	0.22
	(84.78)	(0.75)	(0.50)	(0.68)	(13.49)	(0.36)	(0.09)	(0.46)	(0.41)
Thurgau	560.47	0.09	0.49	-0.05	86.11	0.04	0.65	0.37	0.08
	(75.15)	(0.77)	(0.50)	(0.63)	(12.79)	(0.35)	(0.07)	(0.48)	(0.28)
Zurich	560.72	0.30	0.51	-0.10	78.47	0.19	0.62	0.30	0.26
	(80.00)	(0.77)	(0.50)	(0.61)	(17.36)	(0.43)	(0.10)	(0.46)	(0.44)
Wallis	551.64	0.06	0.49	0.01	89.23	0.02	0.63	0.30	0.43
	(74.77)	(0.74)	(0.50)	(0.43)	(8.51)	(0.31)	(0.10)	(0.46)	(0.50)
Southeastern German Länder	535.01	0.38	0.50	0.03	91.52	0.36	0.62	0.18	0.36
	(80.30)	(0.78)	(0.50)	(0.61)	(8.71)	(0.41)	(0.08)	(0.34)	(0.48)
German city-Länder	507.62	0.40	0.47	0.07	73.80	0.31	0.64	0.15	0.34
	(91.08)	(0.83)	(0.50)	(0.65)	(17.31)	(0.51)	(0.08)	(0.35)	(0.47)
Southwestern German Länder	524.42	0.50	0.51	0.06	81.60	0.44	0.62	0.23	0.36
	(84.40)	(0.78)	(0.50)	(0.60)	(13.74)	(0.43)	(0.09)	(0.42)	(0.48)
Northwestern German Länder	515.71	0.48	0.50	0.05	80.98	0.41	0.64	0.18	0.33
	(84.36)	(0.80)	(0.50)	(0.64)	(12.89)	(0.43)	(0.07)	(0.38)	(0.47)
Northeastern German Länder	509.69	0.32	0.48	0.01	96.49	0.32	0.63	0.05	0.34
	(85.00)	(0.79)	(0.50)	(0.66)	(4.69)	(0.39)	(0.08)	(0.17)	(0.47)
Austria	519.94	0.27	0.50	0.04	90.40	0.22	0.64	0.24	0.24
	(86.63)	(0.78)	(0.50)	(0.55)	(10.57)	(0.39)	(0.07)	(0.43)	(0.43)
Flanders	553.32	0.27	0.46	0.03	94.69	0.25	0.66	0.22	0.48
	(89.47)	(0.85)	(0.50)	(0.47)	(8.95)	(0.46)	(0.08)	(0.41)	(0.50)
Wallonia	518.95	0.27	0.49	0.09	85.84	0.24	0.66	0.15	0.59
	(92.02)	(0.86)	(0.50)	(0.59)	(15.02)	(0.47)	(0.07)	(0.36)	(0.49)
The Netherlands	544.90	0.38	0.49	0.03	91.34	0.32	0.67	0.53	0.24
	(81.54)	(0.82)	(0.50)	(0.56)	(10.90)	(0.40)	(0.06)	(0.50)	(0.43)

Table 2 (continued)

	Low Selection	High Selection	Private Government Independent	Private Government Dependent	School Size	School in Rural Area	School in the City	% Lower Vocational Education	Age of Selection	Central Exit Exam	Socio Economic gradient
Aargau	0.03 (0.16)	0.83 (0.38)	0.00 (0.00)	0.00 (0.00)	354.1 (179.5)	0.96 (0.21)	0.00 (0.00)	15.2	11.00	0	4.51
Bern (German)	0.25 (0.44)	0.47 (0.50)	0.04 (0.19)	0.00 (0.00)	320.1 (216.5)	0.67 (-0.47)	0.01 (-0.30)	34.3	12.00	0	9.81
Basel-Landschaft	0.09 (0.29)	0.61 (0.49)	0.00 (0.00)	0.00 (0.00)	586.5 (294.9)	0.79 (0.41)	0.00 (0.00)	26.5	11.00	0	9.91
St. Gallen	0.23 (0.42)	0.51 (0.50)	0.00 (0.00)	0.00 (0.00)	353.7 (237.5)	0.72 (0.45)	0.00 (0.00)	30.3	12.00	0	6.94
Schaffhausen	0.08 (0.27)	0.53 (0.50)	0.00 (0.00)	0.00 (0.00)	321.8 (240.3)	0.41 (0.49)	0.00 (0.00)	31.1	12.00	0	7.19
Thurgau	0.24 (0.43)	0.38 (0.49)	0.00 (0.00)	0.01 (0.09)	254.4 (172.4)	0.89 (0.31)	0.00 (0.00)	36.6	12.00	0	9.34
Zurich	0.19 (0.39)	0.43 (0.50)	0.00 (0.00)	0.05 (0.22)	409.0 (268.9)	0.51 (0.50)	0.21 (0.41)	29.7	12.00	0	3.57
Wallis	0.35 (0.48)	0.40 (0.49)	0.00 (0.00)	0.00 (0.00)	795.3 (478.7)	0.69 (0.46)	0.00 (0.00)	30.3	12.00	0	6.81
Southeastern German Länder	0.24 (0.42)	0.36 (0.48)	0.00 (0.03)	0.08 (0.26)	532.9 (281.6)	0.45 (0.50)	0.13 (0.34)	18.0	10.00	1	8.90
German city-Länder	0.27 (0.43)	0.09 (0.29)	0.00 (0.02)	0.17 (0.37)	627.7 (284.1)	0.01 (0.05)	0.88 (0.32)	15.0	10.67	0	11.11
Southwestern German Länder	0.19 (0.38)	0.29 (0.43)	0.01 (0.07)	0.13 (0.33)	718.1 (304.0)	0.45 (0.49)	0.12 (0.32)	23.0	10.00	0.7	11.11
Northwestern German Länder	0.16 (0.35)	0.15 (0.35)	0.01 (0.08)	0.10 (0.30)	786.3 (390.4)	0.38 (0.47)	0.18 (0.37)	18.0	10.00	0	12.08
Northeastern German Länder	0.25 (0.42)	0.25 (0.41)	0.00 (0.00)	0.08 (0.27)	497.7 (254.3)	0.45 (0.50)	0.12 (0.31)	5.0	10.00	0.7	8.14
Austria	0.22 (0.41)	0.50 (0.50)	0.00 (0.00)	0.08 (0.27)	547.9 (415.1)	0.48 (0.50)	0.28 (0.45)	23.8	10.00	0	11.24
Flanders	0.31 (0.46)	0.15 (0.35)	0.07 (0.25)	0.69 (0.46)	661.6 (298.2)	0.30 (0.46)	0.11 (0.31)	21.5	12.00	0	9.80
Wallonia	0.37 (0.48)	0.06 (0.24)	0.02 (0.15)	0.59 (0.49)	796.5 (284.6)	0.29 (0.46)	0.21 (0.41)	15.1	12.00	0	3.88
The Netherlands	0.01 (0.11)	0.55 (0.50)	0.00 (0.00)	0.69 (0.46)	1062.8 (538.1)	0.19 (0.39)	0.21 (0.41)	53.3	12.00	1.0	8.60

Source: PISA 2006. The Swiss canton data are from the EDK, the German Länder data are from the IQB, and the data for all the other countries are from the PISA website. The values for the five groups of German Länder are based on observed Länder averages, which we are not allowed to publish by contract with IQB. The southeastern German Länder are Sachsen, Thüringen, and Bavaria; the German city-Länder are Bremen, Hamburg, and Berlin; the southwestern German Länder are Baden-Württemberg, Rheinland-Pfalz, and Saarland; the northwestern German Länder are Hessen, Nordrhein-Westfalen, Niedersachsen, and Schleswig-Holstein; and the northeastern German Länder are Sachsen-Anholt, Brandenburg, and Mecklenburg-Vorpommern.

Table 3: Average parental ESCS level by curriculum in each educational system

	Low	Medium	High
Aargau	0.27	0.37	0.52
Bern (German)	0.08	0.26	0.37
Basel-Landschaft	-0.12	0.08	0.27
St. Gallen	0.02	0.19	0.26
Schaffhausen	-0.07	0.31	0.73
Thurgau	-0.20	0.02	0.08
Zurich	-0.07	0.32	0.68
Wallis	-0.24	0.08	0.19
Southeastern German Länder	-0.18	0.23	0.80
German city-Länder	-0.23	0.21	0.94
Southwestern German Länder	0.01	0.37	0.92
Northwestern German Länder	-0.11	0.36	0.96
Northeastern German Länder	-0.22	0.11	0.77
Austria	-0.02	0.15	0.84
Flanders	-0.33	0.13	0.64
Wallonia	-0.36	0.04	0.53
The Netherlands	0.10	0.55	0.83

Source: PISA 2006. The Swiss canton data are from the EDK, the German Länder data are from the IQB, and the data for all the other countries are from the PISA website. The averages for the five groups of German Länder are based on observed Länder values, which we are not allowed to publish by contract with IQB. The southeastern German Länder are Sachsen, Thüringen, and Bavaria; the German city-Länder are Bremen, Hamburg, and Berlin; the southwestern German Länder are Baden-Württemberg, Rheinland-Pfalz, and Saarland; the northwestern German Länder are Hessen, Nordrhein-Westfalen, Niedersachsen, and Schleswig-Holstein; and the northeastern German Länder are Sachsen-Anhalt; Brandenburg, and Mecklenburg-Vorpommern.

Table 4: Coefficients for Flanders and the Netherlands in a series of multilevel analyses

	Flanders	Netherlands
M1: no controls	40.9** (14.0)	50.1** (25.1)
M2: controls for all individual characteristics	38.8** (14.2)	38.9 (24.6)
M3: M2 & % natives, average school ESCS, ESCS school diversity	34.6* (13.8)	18.5 (31.0)
M4: M2 & entrance selection school	25.3 (15.5)	22.2 (15.5)
M5: M2 & school size, school size squared	45.7 (24.3)	10.1 (39.9)
M6: M2 & curriculum level	45.8* (20.8)	39.2 (34.9)
M7: M2 & controls for all curriculum characteristics	41.7* (19.0)	22.1 (35.2)
M8: M2 & central exit exams	40.1** (12.8)	29.3 (25.4)
M9: M2 & age of selection	39.1** (14.5)	39.9 (25.7)
M10: M2 & socioeconomic gradient	41.0** (14.4)	39.1 (24.7)
M11: M2 & % lower vocational education	36.6* (14.3)	30.0 (24.5)
M24: all individual, curriculum & system controls	56.2** (18.4)	14.5 (25.9)

Source: See appendices II to V for full equations.

Table 5: Multilevel analysis of the determinants of math scores (individual, curriculum, and educational system levels), with $N_i = 48385$; $N_s = 2816$; $N_{es} = 26$; $N_{ns} = 5$.

	Model 1	Model 2	Model 7	Model 24
Flanders	40.9** (14.0)	38.8** (14.2)	41.7* (19.0)	56.2** (18.4)
The Netherlands	50.1* (25.1)	38.9 (24.6)	22.1 (35.2)	14.51 (25.9)
Individual				
Parental ESCS		11.1*** (0.4)	7.2*** (0.4)	7.5*** (0.4)
Female		-26.0*** (0.5)	-27.5*** (0.6)	-27.5*** (0.5)
Grade		27.8*** (0.5)	24.7*** (0.5)	25.6*** (0.4)
Curriculum				
Average school ESCS			26.0*** (2.0)	26.3*** (1.9)
ESCS school diversity			10.9 (7.3)	7.5 (7.0)
% Natives			0.15** (0.05)	0.15** (0.04)
Lower vocational curriculum (ref. medium)			-57.6*** (1.4)	-56.8*** (1.4)
High curriculum (ref. medium)			54.4*** (1.5)	55.0*** (1.4)
Low selection (ref. medium selection)			-4.4** (1.5)	-3.6** (1.4)
High selection (ref. medium selection)			7.3*** (1.4)	7.8*** (1.3)
Private gov. independent (ref. public)			-16.4** (5.3)	-18.5*** (5.2)
Private gov. dependent (ref. public)			8.3*** (1.7)	3.2* (1.5)
School size			0.03*** (0.01)	0.03*** (0.00)
School size squared			-0.00001** (0.00)	-0.00001** (0.00)
Rural area (ref. town)			-8.9*** (1.7)	1.8 (1.3)
City (ref. town)			0.1 (0.1)	-6.8*** (1.6)
Educational system				
Central exit exam				9.5 (6.4)
Age of selection				11.1* (4.5)
Socioeconomic gradient				-0.4 (0.6)
% lower vocational education				1.13*** (0.29)
Constant	509.2*** (11.2)	522.5*** (10.9)	489.5*** (18.0)	277.6*** (58.6)
Variance components				
National state level	446.6*** (405.1)	425.0*** (376.8)	1029.7*** (861.9)	126.0** (205.9)
Educational system level	87.1*** (40.7)	95.9*** (41.0)	200.3*** (65.0)	130.4*** (48.7)
Curriculum level	4131.8*** (118.7)	3214.3*** (94.9)	688.3*** (26.1)	688.9*** (24.4)
Individual level	3407.3*** (22.5)	2982.2*** (19.8)	2817.3*** (19.9)	2809.0*** (18.6)

Source: PISA 2006. The Swiss canton data are from the EDK, the German Länder data are from the IQB, and the data for all the other countries are from the PISA website.

Appendix I: Definition of curriculum levels

We classify the curricula into three different categories: 1) low for low vocational education; 2) medium for higher vocational education and lower general education, with no access to tertiary education; and 3) high for higher general education that gives direct access to tertiary education upon finishing high school. We use the original version of the so-called unique national study programme code from the PISA 2006 codebook. Table A1 illustrates the categorization of the curricula in each of the countries considered.

Table A1: curriculum level per country.

Country/ curriculum	Low	Medium	High
Austria	Vocational program	Lower secondary school	Gymnasium lower secondary
	Middle vocational school	Upper secondary school	Gymnasium upper secondary
	Vocational college	Apprenticeship	
		Higher vocational school	
	Second year of first stage—preparing for vocational secondary education	Second- & third-stage technical secondary education	(First year a of first stage of) general education
2nd & 3rd vocational secondary education	Special sec. educ. - upper sec. (training form 3 / years 4 and 5)	Second year of first stage preparing for regular sec. educ.	
Part-time vocational secondary education focused on the labor markets	2nd & 3rd degrees of techn. or art. educ. (transition) (FR/GER)	Second & third stage regular secondary education	
2nd year of vocational education (French/German)	2nd & 3rd degrees of techn. or art. Educ. (qualif.) (FR/GER)	First degree of general education (FR/GER)	
Belgium	Complementary year to first degree (French community only)	Special sec. Educ. (upper sec.-training form 3) (FR only)	Second & third degrees of general education (FR/GER)
	2nd & 3rd degrees of vocational education (FR/GER)		
	Vocational training focused on the labor market (FR com only)		
	Special sec. educ. (lower sec. - training form 3) (Wallonia only)		
	Special sec. educ. (lower sec.) (GER only)		
Germany	Lower sec., no access to upper secondary (<i>Hauptschule</i>)	Lower secondary with access to upper secondary (comprehensive)	Lower sec. With access to upper secondary (Gymnasium)
	Lower sec., no access to upper sec. (<i>koop. Gesamtschule Hauptschule</i>)	Lower sec., no access to upper secondary (<i>Realschule</i>)	Upper sec. Level (Gymnasium)
	Lower sec., no access to upper secondary (<i>Hauptschule</i> integrated)	Comprehensive lower sec. with access to upper sec. (<i>Gemamtschule</i>)	Lower sec., with access to upper sec. (Koop. Gs, Gymn.)
	Lower sec., no access to upper secondary (<i>Hauptschulklasse</i>)	Lower sec., with or without access to upper sec. (koop. Gesamtschules, Realschule)	Lower secondary with access to upper secondary (Waldorf)
	Pre-vocational training year	Lower sec., no access to upper secondary (<i>Realschule</i>)	Upper secondary level of education (Waldorf)

	Vocational school (<i>Berufsfachschule</i>)	integrated) Lower sec., no access to upper secondary (<i>Realschulklasse</i>) Vocational school (<i>Berufsschule</i>)	
	Practical preparation for labor market Vmbo (general voc.) Vmbo bb (1-2 year) Vmbo bb (3-4 year) Vmbo kb (1-2 year) Vmbo kb (3-4 year) Vmbo gl/ tl (1-2 year) Vmbo gl/ tl (3-4 year)	Havo (year 1-3) Havo (sec. year 4-5)	Vwo (year 1-3) Vwo (year 4-6)
The Netherlands			
Switzerland ²⁴ (German cantons)	Basic requirements Heterogeneous basic requirements	Intermediate requirements Heterogeneous intermediate requirements	Higher requirements Heterogeneous higher requirements

Source: PISA 2006 codebook.

²⁴ Pupils are allocated to institutionally separate school types, according to their performance levels. The structure is based on the principle of equal capacities among pupils. Generally, each school type has its own adapted curricula, teaching material, and teachers and, in some cases, its own range of subjects. In general, there are two to three school types (four in a minority of cantons), whose names vary. In the structure with two school types, a distinction is made between the performance-based group at the basic level (with the least demanding requirements) and the performance-based group at the advanced level. In the structure with three school types, there is a performance-based group at the basic, intermediate, and advanced levels. The requirements of the performance-based group at the advanced level are the most demanding and this school type generally prepares pupils for transfer to the Matura schools.

Appendix II: Multilevel analysis of the determinants of math scores (individual and curriculum controls), with $N_i = 48385$, $N_s = 2816$, $N_{es} = 26$, $N_{ns} = 5$

	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7
Flanders	40.92**	38.78**	34.61*	25.27	45.67	45.80*	41.73*
	(14.030)	(14.210)	(13.760)	(15.470)	(24.310)	(20.800)	(19.040)
Netherlands	50.06*	38.93	18.49	22.24	10.06	39.15	22.07
	(25.060)	(24.550)	(31.010)	(15.520)	(39.890)	(34.940)	(35.200)
Individual level							
Parental ESCS		11.09***	9.233***	10.80***	10.38***	8.260***	7.247***
		(0.362)	(0.365)	(0.362)	(0.360)	(0.353)	(0.379)
Female		-26.01***	-26.21***	-26.28***	-25.93***	-27.58***	-27.52***
		(0.534)	(0.531)	(0.532)	(0.529)	(0.515)	(0.551)
Grade		27.82***	27.50***	27.68***	27.23***	25.81***	24.68***
		(0.458)	(0.455)	(0.456)	(0.454)	(0.441)	(0.474)
Curriculum							
% natives			0.114*				0.153**
			(0.051)				(0.047)
Average school ESCS			95.57***				25.96***
			(1.855)				(1.966)
ESCS school diversity			45.43***				10.91
			(8.819)				(7.296)
Low selection				-20.82***			-4.404**
(ref. medium selection)				(2.147)			(1.464)
High selection				31.49***			7.332***
(ref. medium selection)				(1.939)			(1.388)
School size					0.170***		0.0256***
					(0.006)		(0.005)
School size squared					-0.0000575**		-0.00000704*
					(0.000)		(0.000)
Lower vocational curriculum						-67.90***	-57.55***
(ref. medium)						(1.322)	(1.444)
High						68.89***	54.35***

curriculum							
(ref. medium)						(1.265)	(1.473)
Private gov. indep.							-16.36**
(ref. public)							(5.349)
Private gov. dep.							8.303***
(ref. public)							(1.701)
City							-8.878***
(ref. town)							(1.703)
Rural							0.124
(ref. town)							(0.09)
Constant	509.2***	522.5***	472.1***	521.4***	452.8***	527.0***	489.5***
	(11.170)	(10.930)	(16.230)	(3.309)	(19.090)	(16.270)	(18.030)
Variance components							
National state level	446.6***	425.0***	797.9***	x	1325.6***	967.3***	1029.7***
	(405.100)	(376.800)	(649.200)		(1140.100)	(835.500)	(861.900)
Educational system level	87.07***	95.94***	93.96***	223.9***	338.6***	247.1***	200.3***
	(40.740)	(40.980)	(34.290)	(77.220)	(114.300)	(78.960)	(64.990)
Curriculum level	4131.8**	3214.3**	1264.1**	2995.9**	2630.9***	820.1***	688.3***
	(118.700)	(94.890)	(40.980)	(89.080)	(79.010)	(27.910)	(26.140)
Individual level	3407.3**	2982.2**	2983.0**	2964.4**	2935.5***	2815.7**	2817.3***
	(22.480)	(19.770)	(19.770)	(19.650)	(19.460)	(18.640)	(19.900)

Source: PISA 2006. The Swiss canton data are from the EDK, the German Länder data are from the IQB, and all the data for the other countries are from the PISA website. Here the x = four-level model did not converge, so the three-level model is used.

Appendix III: Multilevel analysis of the determinants of math scores (individual and educational system controls), with $N_i = 48385$, $N_s = 2816$, $N_{es} = 26$, $N_{ns} = 5$

	Model 8	Model 9	Model 10	Model 11
Netherlands	29.27	39.86	39.09	29.98
	(25.440)	(25.680)	(24.740)	(24.470)
Flanders	40.14**	39.11**	41.03**	36.61*
	(12.810)	(14.540)	(14.410)	(14.290)
<i>Individual level</i>				
Parental ESCS	11.10***	11.09***	11.09***	11.09***
	(0.362)	(0.362)	(0.362)	(0.362)
Female	-26.01***	-26.01***	-26.01***	-26.01***
	(0.534)	(0.534)	(0.534)	(0.534)
Grade	27.82***	27.82***	27.82***	27.82***
	(0.458)	(0.458)	(0.458)	(0.458)
<i>Educational system</i>				
Central exit exam	12.34*			
	(5.234)			
Age of selection		-0.89		
		(4.796)		
Socioeconomic gradient			-0.446	
			(0.491)	
% lower vocational education				0.291
				(0.278)
Constant	520.6***	532.2***	526.1***	516.4***
	(11.340)	(53.850)	(11.760)	(11.770)
<i>Variance components</i>				
National state level	471.9***	455.8***	434.3***	355.4***
	(404.400)	(406.200)	(387.500)	(330.200)
Educational system level	72.11***	100.7***	95.80***	98.13***
	(34.230)	(43.340)	(42.300)	(42.640)
Curriculum level	3213.3***	3214.4***	3214.4***	3214.3***
	(94.860)	(94.900)	(94.900)	(94.900)
Individual level	2982.3***	2982.2***	2982.2***	2982.2***
	(19.770)	(19.770)	(19.770)	(19.770)

Source: PISA 2006. The Swiss canton data are from the EDK, the German Länder data are from the IQB, and the data for all the other countries are from the PISA website.

Appendix IV: Multilevel analysis of the determinants of math scores (individual, entrance selection, and educational system controls), with $N_i = 48385$, $N_s = 2816$, $N_{es} = 26$, $N_{ns} = 5$

	Model 12	Model 13	Model 14	Model 15
Netherlands	24.06	8.533	21.63	-6.702
	-(16.670)	-(13.590)	-(15.210)	-(15.480)
Flanders	24.61	11.56	25.72	23.74
	-(15.940)	-(13.540)	-(15.160)	-(12.840)
Individual level				
Parental ESCS	10.80***	10.81***	10.80***	10.80***
	-(0.362)	-(0.362)	-(0.362)	-(0.362)
Female	-26.28***	-26.28***	-26.28***	-26.28***
	-(0.532)	-(0.532)	-(0.532)	-(0.532)
Grade	27.68***	27.68***	27.68***	27.68***
	-(0.456)	-(0.456)	-(0.456)	-(0.456)
Educational system				
Central exit exam	-2.557			
	-(7.296)			
Age of selection		10.33***		
		-(3.021)		
Socioeconomic gradient			-0.911	
			-(0.630)	
% lower vocational education				0.884***
				-(0.263)
School				
Low selection	-20.81***	-20.91***	-20.90***	-20.85***
(ref. medium selection)	-(2.147)	-(2.145)	-(2.147)	-(2.145)
High selection	31.47***	31.44***	31.42***	31.42***
(ref. medium selection)	-(1.939)	-(1.936)	-(1.940)	-(1.937)
Constant	522.2***	411.3***	529.9***	503.7***
	-(3.922)	-(32.270)	-(6.710)	-(5.923)
Variance components				
National state level	x	x	x	x
Educational system level	235.2***	150.4***	213.9***	147.5***
	-(81.890)	-(53.750)	-(75.670)	-(54.680)
Curriculum/school level	2995.7***	2993.2***	2995.6***	2994.9***
	-(89.070)	-(88.950)	-(89.060)	-(89.020)
Individual level	2964.4***	2964.4***	2964.4***	2964.4***
	-(19.650)	-(19.650)	-(19.650)	-(19.650)

Source: PISA 2006. The Swiss canton data are from the EDK, the German Länder data are from the IQB, and the data for all the other countries are from the PISA website. The x = four-level model did not converge, so the three-level model is used.

High selection									7.791***
(ref. medium selection)									(1.329)
Private gov. independent									-18.49***
(ref. public)									(5.163)
Private gov. dependent									3.155*
(ref. public)									(1.519)
School size									0.0249***
									(0.004)
School size squared									-0.00000620**
									(0.000)
Rural area									1.814
(ref. town)									(1.313)
City									-6.837***
(ref. town)									(1.563)
Education system									
Central exit exam	11.41*				13.65				9.455
	(4.753)				(7.646)				(6.408)
Age of selection		4.011				3.654			11.11*
		(4.485)				(6.653)			(4.526)
Socioeconomic gradient			-0.231				-0.528		-0.42
			(0.456)				(0.688)		(0.583)
% lower vocational education				0.288				1.162**	1.139***
				(0.252)				(0.309)	(0.288)
Constant	471.4**	428.1**	474.1**	465.4**	524.8**	486.9**	531.5**	501.3**	277.6***
	(16.690)	(51.780)	(16.650)	(16.670)	(16.750)	(74.970)	(17.370)	(14.550)	(58.600)
Variance components									

National state level	873.9** *	723.0** *	793.0** *	711.7** *	1040.0 ***	967.0** *	983.4** *	596.1** *	126.0**
	(705.8 00)	(604.6 00)	(646.7 00)	(589.1 00)	(875.4 00)	(854.3 00)	(855.1 00)	(533.4 00)	(205.900)
Educational system level	72.00** *	97.53** *	98.06** *	93.57** *	222.1** *	256.5** *	251.2** *	158.9** *	130.4***
	(28.64 0)	(35.88 0)	(36.39 0)	(35.00 0)	(73.25 0)	(83.33 0)	(82.25 0)	(54.23 0)	(48.650)
Curriculum level	1263.9 ***	1264.1 ***	1264.1 ***	1264.1 ***	820.3** *	820.1** *	820.1** *	820.1** *	688.9***
	(40.97 0)	(40.98 0)	(40.98 0)	(40.98 0)	(27.91 0)	(27.90 0)	(27.91 0)	(27.91 0)	(24.350)
Individual level	2983.0 ***	2982.9 ***	2983.0 ***	2983.0 ***	2815.6 ***	2815.7 ***	2815.7 ***	2815.7 ***	2809.0***
	(19.77 0)	(19.77 0)	(19.77 0)	(19.77 0)	(18.64 0)	(18.64 0)	(18.64 0)	(18.64 0)	(18.600)

Source: PISA 2006. The Swiss canton data are from the EDK, the German Länder data are from the IQB, and the data for all the other countries are from the PISA website.