



**BGPE Discussion Paper**

**No. 99**

**Are Immigrants and Girls Graded Worse?  
Results of a Matching Approach**

**David Kiss**

May 2011

---

ISSN 1863-5733

Editor: Prof. Regina T. Riphahn, Ph.D.  
Friedrich-Alexander-University Erlangen-Nuremberg  
© David Kiss

---

# Are Immigrants and Girls Graded Worse? Results of a Matching Approach

David Kiss\*, April 01, 2011

## Abstract:

Using PIRLS 2001 and PISA 2003 data for Germany, this paper examines whether second-generation immigrants and girls are graded worse in math than comparable natives and boys, respectively. Once all grading-relevant characteristics, namely math skills and oral participation, are accounted for, pupils should obtain same school grades. Results of a matching approach and class fixed effects regressions suggest that second-generation immigrants have grade disadvantages in primary education which could bias their secondary school track choice. Regarding secondary school, most immigrants are not affected by grade discrimination and girls enrolled in upper-secondary school are systematically graded better. *JEL: C40, I21, J15.*

*Keywords: grade discrimination in primary and secondary education, school careers of immigrants and girls, matching frameworks*

---

\* University of Erlangen-Nuremberg, Lange Gasse 20, D-90403 Nuremberg, Germany, david.kiss@wiso.uni-erlangen.de (former affiliation: RWI, Hohenzollernstr. 1-3, D-45128 Essen, Germany). I am grateful to Guido Schwerdt for motivating my research on educational outcomes of immigrants. Many thanks to Leilanie Basilio, Thomas K. Bauer, Alfredo R. Paloyo, Regina T. Riphahn, Marcus Tamm, and participants of the International Workshop on Applied Economics of Education (IWAE 2010) and the Spring Meeting of Young Economists 2011 (SMYE 2011) for valuable discussions and suggestions. This paper largely benefited from the comments of two anonymous referees which is gratefully acknowledged. I would also like to thank the WGL ("Initiative for Research", project leader: Andrea Weber) for financial support. The data were produced by IFS Dortmund (PIRLS 2001) and IPN Kiel (PISA 2003) and provided to me by the research data center (FDZ) at IQB Berlin, whose staff I thank for their help and support. All remaining errors are my own.

## 1. Introduction

School grades largely determine future labor market success for several reasons. In Germany, transition into different secondary school types mainly depends on ability as measured by the average of grades in math and German. Hence grade discrimination at the end of primary education can bias track choice. As shown in Dustmann (2004), students attending higher secondary school tracks typically accumulate more human capital which translates into higher earnings. Final grades and degrees from secondary school are also relevant for hiring decisions of employers or access to university subjects with limited admission.<sup>1</sup>

Using two representative cross-sectional surveys for German primary and secondary schools, "Progress in International Reading Literacy Study 2001" (PIRLS 2001) and "Programme for International Student Assessment 2003" (PISA 2003), this paper examines whether second-generation immigrants and girls are graded worse in math than comparable natives or boys, respectively. Pupils should obtain same grades once differences in grading-relevant characteristics, namely math test scores and oral participation, are accounted for. Many studies that investigate grade discrimination make use of experimental settings. In van Ewijk (2010), typical Dutch, Turkish, and Moroccan names are assigned to identical sets of essays written by 11-year old pupils, and total number of 113 Dutch teachers are asked to grade them. Since immigrants and natives are graded similarly, no evidence for grade discrimination in Dutch primary schools is given.<sup>2</sup> Essays written by Swedish high school pupils are subject to blind and non-blind grading in Hinnerich et al. (2011). They provide evidence for grade discrimination against immigrant students because non-blind graded natives obtain bet-

---

<sup>1</sup> Unlike human capital theory, the large literature on signalling and screening (starting with Spence (1973) and Arrow (1973)) argues that educational attainment reveals unobserved characteristics like innate ability. Empirical evidence for statistical discrimination on the basis of education is given in, for instance, Altonji and Pierret (2001).

<sup>2</sup> However, van Ewijk (2010) finds that ethnic minority students are more likely to be considered as low-performers with unfavorable attitudes which might reduce their learning effort. The impact of teacher expectations on student behavior is also examined in Dee (2005): using a representative survey of US-American 8th-graders, he finds that black students are more often considered as inattentive and disruptive if their teacher was from a different ethnic background.

ter grades than non-blind immigrants. At the same time, however, grade differentials between blind graded natives and immigrants are insignificant.

Regarding possible grade discrimination among boys and girls, the same data are examined in Hinnerich et al. (2010). Blind and non-blind graded boys obtain lower grades than girls but grade differentials between boys and girls are similar in the blind and non-blind case. Lavy (2008) compares blind and non-blind grades of Israeli high school graduates. Unlike Hinnerich et al. (2010), he finds evidence for discrimination against boys in various subjects. To my knowledge, no studies investigate grade discrimination between German boys and girls.

Only a small number of studies examined for German schools whether immigrant pupils are graded worse than natives once differences in grading-relevant characteristics are controlled for. Using PIRLS 2001 data, which were restricted to Western Germany, Luedemann and Schwerdt (2010) find that second-generation immigrants are less likely to receive teacher recommendations for higher secondary school tracks. Further, OLS estimates indicate that second-generation immigrants are graded worse in math at the end of primary education. Sprietsma (2009) also provides evidence for grade discrimination in primary education: similar to van Ewijk (2010), typical German and Turkish names are randomly assigned to same sets of essays written by fourth-graders. She finds that essays bearing Turkish names are graded significantly worse.

The analysis of the data leads to two main findings: (i) results of a matching approach and class fixed effects estimates suggest that second-generation immigrants attending primary school obtain worse grades than comparable natives. After track choice most second-generation immigrants are not graded differently from natives. The more precise estimates of the matching approach, however, indicate that averagely performing second-generation immigrants enrolled in middle-secondary school might suffer from grade discrimination. (ii) No evidence for grade discrimination against native girls is given here, except girls in upper-

secondary school. Once differences in grading-relevant characteristics are accounted for, class fixed effects and matching estimates show that girls in upper-secondary school obtain better grades than comparable boys.

This study contributes to the literature in several ways: (i) provided that math test scores and (self-reported) oral participation sufficiently capture a pupil's grading-relevant characteristics in math, results should not be biased because of omitted variables.<sup>3</sup> (ii) To detect grade discrimination patterns, a nonparametric matching approach from the labor market literature is suggested. It provides more reliable and precise estimates than class fixed effects regressions because pupils from the treatment and control groups are only taken into account if they attend the same class, have similar test scores, and similar levels of oral participation. Unlike OLS, matching does not assume a linear relationship between grades and test scores, which is a strong assumption.

The paper is organized as follows. Section 2 briefly describes the data and institutional background. The matching approach and class fixed effects model are presented in section 3. Section 4 reports the results. The difference between detection and explanation of discrimination is discussed in section 5. Section 6 concludes.

## **2. Data and Institutional Background**

In Germany, primary education generally ends after the fourth grade when children are 10 or 11 years old. Thereafter students are tracked by ability into three different types of secondary schools: lower-secondary (Hauptschule), middle-secondary (Realschule), and upper-secondary school (Gymnasium). Upper-secondary school is the most academic track and pre-

---

<sup>3</sup> For instance, the data do not contain an objective measure for spoken language skills. Thus estimating the impact of migration background on grades in German is potentially biased if differences in reading skills are solely accounted for.

pares students for university study. Middle- and lower-secondary school pupils usually start vocational training after completion of their secondary education.<sup>4</sup>

Regarding the data used in this study, PIRLS 2001 focuses on pupils at the end of their primary education while PISA 2003 covers ninth-graders in secondary school.<sup>5</sup> Math tests used multiple choice and open-ended questions. Figure 1a and Figure 1b show some example exercises from PIRLS and PISA, respectively. These exercises also control for language skills because pupils with a bad command of the German language may not fully understand the exercises which would lead to lower test scores. Apart from test scores, additional data were collected from questionnaires completed by students, parents, teachers and principals. Most information about the pupil, namely his/her math grades in the last school report, self-reported oral participation, immigrant background, and sex are obtained from the student questionnaire.<sup>6</sup> In both surveys, schools are the primary sampling unit. Within each school, classes are randomly chosen. All pupils attending a drawn class participated in the tests on a voluntary basis. In this study, observations with missing values in math grades, test scores, and oral participation are deleted. The samples are additionally restricted on pupils with known migration background and sex.

Documents from "the Standing Conference of the Ministers of Education and Cultural Affairs of the Laender in the Federal Republic of Germany", which decides nationwide educational standards, state that grading in the subjects math (and German) should mainly depend on the student's performance in written exams and his/her oral participation (Kultusministerkonferenz (2009): 130). Therefore students with same skills and levels of oral participation should not be graded differently. Table 1 contains summary statistics for differ-

---

<sup>4</sup> There are additional secondary school types, like comprehensive schools (Gesamtschulen). These are left out from the analysis because school-leaving certificates from these schools cannot be clearly ranked with certificates from upper-, middle-, or lower-secondary school. In PISA 2003, 20% of the pupils attend alternative secondary school tracks.

<sup>5</sup> PISA 2003 consists of three parts: PISA-international (PISA-I) is the smallest data set. It was collected for international comparisons. In Germany, this data set has been extended twice: PISA-I-PLUS (which is used here) and PISA-E.

<sup>6</sup> PIRLS took place in May 2001 while PISA was conducted from March to August 2003. Pupils receive their half-yearly school reports by the end of February.

ent groups of pupils, separated by school types. For the purpose of this study, math grades range from 0 (fail) to 4 (very good). The higher a pupil's test score from the standardized math test, the higher his/her math proficiency. Within a school type, math test scores have mean 0 and standard deviation 1. Unlike math test scores, the data do not contain an objective measure for oral participation. However, pupils were asked to which extent they agree with the statement "I frequently participate in math lessons". The four outcomes of this categorical variable are "I completely disagree", "I somewhat disagree", "I somewhat agree" and "I completely agree". Here, that variable ranges from 1 (very low) to 4 (very high). To facilitate comparisons, it is not turned into four indicator variables in the summary statistics.

Columns 1-3 in Table 1 report summary statistics for natives and second-generation immigrants enrolled at different school types. In this study, second (or higher) generation immigrants are born in Germany and satisfy at least one of the following conditions: (i) at least one parent was born abroad or (ii) languages other than German are (also) spoken at home. Pupils are considered as natives, if they are born in Germany, as well as their parents, and German is the only language spoken at home. A comparison between columns 1 and 2 suggests that natives obtain better math grades than second-generation immigrants. For instance, natives in primary education are significantly graded better than second-generation immigrants (2.53 vs. 2.11,  $p$ -value  $< 0.01$ ). Natives also obtain better grades in upper-secondary school (2.02. vs. 1.92). However, the  $p$ -value of 0.35 indicates that the difference in means is not significant. In the primary education sample, a total number of 1701 pupils are enrolled in 261 classes (column 3). Among these pupils, 354 pupils from 179 classes are second-generation immigrants (column 2). Regarding the grading-relevant characteristic "math test score", second-generation immigrants are outperformed by natives.<sup>7</sup> Second-generation immigrants enrolled in middle- and lower-secondary school report higher levels of oral participation than natives, however, the  $p$ -values indicate that differences in means is not significant.

---

<sup>7</sup> This is consistent with other findings about immigrants' educational attainment in Germany, see Riphahn (2001) or Schnepf (2007).

Regarding native boys and girls, the patterns are similar across school types (columns 4-6): boys have more favorable grading-relevant characteristics and better math grades. Only in upper-secondary school, boys and girls obtain same grades on average. In few cases, information about the pupil's sex is missing. Therefore sample sizes reported in column 6 are slightly smaller than in column 1. For example, 1347 natives attend primary school (column 1). Since one native did not report his/her sex, the number of native boys and girls in primary school is 1346 in column 6.

### 3. Empirical Strategy

#### *Class Fixed Effects Model*

In the following, the term "treatment group" refers to second-generation immigrants or girls. Depending on the chosen treatment group, natives or boys constitute the control group. To keep things simple, second-generation immigrants are considered here as the treatment group. The impact of immigrant background on math grades is then estimated with the following model:

$$y_{ic} = \beta_0 + \beta_1 tsmat_{ic} + \beta_2 freq_{ic} + \gamma SGI_{ic} + \lambda_c + \varepsilon_{ic}. \quad (1)$$

$y_{ic}$  is the self-reported math grade of student  $i$  attending class  $c$ .  $tsmat$  is the math test score. Within a schooling level or track,  $tsmat$  has mean 0 and standard deviation 1. The dummy variable  $freq$  indicates "frequent participation in math lessons". It has value 1 whenever a pupil's level of self-reported oral participation is 3 (high) or 4 (very high).<sup>8</sup> Class fixed effects are picked up by  $\lambda$ , while errors  $\varepsilon$  are clustered on the class level. Immigrant background is captured by the indicator variable  $SGI$ .<sup>9</sup> Provided that all grading-relevant charac-

---

<sup>8</sup> Consequently, "frequent participation" has value 0 whenever pupils report "low" or "very low" levels of oral participation. Estimates in Table 2 and Table 3 are better comparable with results of the matching approach if oral participation is not decomposed into four dummy variables. Estimating model (1) with "oral participation" instead of "frequent participation" does not change the results (estimates available on request).

<sup>9</sup> If boys and girls are compared,  $SGI$  is replaced by the indicator variable  $GIRL$ , and samples are restricted on natives with known sex.



teristics are sufficiently captured in model (1), significant estimates of  $\gamma$  indicate (positive or negative) grade discrimination.

In Germany, subjects are usually taught by the same teacher throughout the whole school year. Hence, class fixed effects  $\lambda$  also control for the average teacher effect on math grades. Grading practices usually differ among teachers. For instance, they can be influenced by average skills on class level (Himmler and Schwager (2007), Dardanoni et al. (2009)), the (controversial) impact of class size and its determinants on students' performance (Angrist and Lavy (1999), Woessmann and West (2006)), or the teacher's aspiration level (Iacus and Porro (2007)).

### *Matching Approach*

To interpret class fixed estimates more precisely, a matching approach from the labor market literature is additionally implemented. Nopo (2008) uses a nonparametric alternative to the classical Blinder-Oaxaca decomposition (see Blinder (1973)) to estimate gender wage gaps.<sup>10</sup> His approach depends on matching men and women who are similar in terms of labor market-relevant characteristics like education or potential work experience and calculating wage differentials. Individuals from the control and treatment groups are directly matched on observed characteristics. For the purpose of this study, the wage gap between men and women is transformed into a math grade gap between immigrants and natives with similar math test scores and levels of oral participation. In the following, the intuition behind the matching approach is outlined for the treatment "second-generation immigrant". Same reasoning applies to the case "girls vs. boys".

Results of the matching approach are obtained from "statistical twin pairs" of matched immigrants and natives. The grade difference within pairs is the variable of interest. The procedure can be split up into two steps. First, each immigrant in the data is matched on a native

---

<sup>10</sup> Nopo (2008) often refers to Barsky et al. (2002) who analyze the black-white wealth gap in the United States.

pupil if three conditions are fulfilled: both pupils attend the same class, have similar math test scores, and similar levels of oral participation.<sup>11</sup> Immigrants are left out if the data contain no "suitable" native child. Matching on the class ID ensures that both children are taught by the same teacher.<sup>12</sup> Sometimes an immigrant child can be matched on more than one native. In these cases,  $T - 1$  "clones" of the immigrant child are created.

In the second step, the grade difference within each statistical twin pair is calculated and weighted by  $T^{-1}$ , the inverse of the number of suitable natives. Math grade gaps are denoted by  $\Delta_{cst}^{MG}$ .<sup>13</sup> By definition, positive values of  $\Delta^{MG}$  indicate grade advantages for immigrants and vice versa. Consequently,  $\Delta^{MG} = 0$  means that the matched immigrant and native are not graded differently.

Compared to OLS, matching has two advantages: (i) OLS assumes a linear relationship between grades and test scores which is a strong assumption. Matching only presupposes that same grading practices are applied on (in terms of grading-relevant characteristics) identical pupils. Among teachers, matching allows different nonparametric relationships between grades and relevant characteristics.<sup>14</sup> (ii) Matching results solely depend on pupils who are comparable in terms of grading-relevant characteristics. In OLS, unmatched immigrants and natives also contribute to the estimation of the parameters. However, the more variables are considered as grading-relevant, the smaller the number of statistical twin pairs.

---

<sup>11</sup> Test scores are continuous. Therefore immigrants are matched on natives who are "slightly" better or worse in terms of test scores. Depending on estimates from model (1), each pupil's math grade is predicted. For the purpose of the matching approach, that prediction solely depends on math test scores. Since math grades range from 0 to 4, differences in predicted math grades smaller than  $|0.1\bar{6}|$  are assumed to indicate similarity in math skills. Pupils are further matched on values of the dummy variable "frequent participation". As already mentioned, "frequent participation" has value 1 whenever a pupil reports "high" or "very high" levels of oral participation (and value 0 whenever a pupil reports "low" or "very low" levels of oral participation).

<sup>12</sup> Matching on the class ID instead of the teacher ID yields more reliable results because pupils might be grouped by ability into classes. Even if ability grouping does not take place, two classes taught by the same teacher could differ, for instance, in their progress in the curriculum. This might cause heterogeneity in applied grading practices.

<sup>13</sup> "MG" means "math grade", the index  $cst$  is the (unique) identifier of twin pairs:  $c$  is the index of the class,  $s \in S_c$  refers to the set of grading-relevant characteristics  $S_c$  that constitute a statistical twin pair in class  $c$ .  $t$  is the index of the (cloned) immigrant.

<sup>14</sup> As shown in Iacus and Porro (2007), the relationship between grades and test scores is not linear, but more complex. It is further shown that the (nonparametric) relationship between grades and test scores strongly varies among teachers and schools.

#### 4. Results

Table 2a and b report class fixed effects estimates of immigrant background and female sex, respectively, on math grades. As expected, the significant and large point estimates for math test scores and frequent participation suggest that both characteristics are positively correlated with better grades.

Compared to natives, second-generation immigrants attending primary school are graded worse 0.27 grades on average, which is substantial. Figure 2 plots the distribution of grade gaps  $\Delta_{cst}^{MG}$  of matched second-generation immigrants who are enrolled in primary school. Here, 152 second-generation immigrants are matched on 219 natives.<sup>15</sup> For test scores smaller than 0.5, the 90% confidence band of the smoothed regression line (kernel-weighted local polynomial smoothing) runs below the zero-grade-gap line, which indicates grade disadvantages for averagely- and low-performing immigrants. Second-generation immigrants with test scores greater than 0.5 are not graded differently from comparable natives because the 90% confidence band overlaps with the zero-grade-gap line.

The results of the matching approach suggest that the negative point estimate for second-generation immigrants in Table 2a is caused by averagely and low-performing second-generation immigrants. However, matching estimates are biased if immigrants are systematically matched on natives with more favorable grading-relevant characteristics. Supplementary summary statistics are therefore added to Figure 2. The three columns refer to different percentiles in the test score distribution of matched pupils. As expected, mean test scores of statistical twin pairs are ascending ( $-0.884 < -0.054 < 0.837$ ).  $\Delta^{TS}$  is the test score difference within twin pairs. For any twin pair  $cst$ ,  $\Delta_{cst}^{TS} < 0$  implies that the matched immigrant has a smaller test score than his/her native "twin brother/sister". Since math test scores are continu-

---

<sup>15</sup> Since the number of matched second-generation immigrants exceeds the number of statistical twin pairs (152 vs. 219), some second-generation immigrants have been "cloned". It is also noteworthy that 152 is 43% of 354, which is the number of second-generation immigrants in primary education (see Table 1, column 2). In the light of the restrictive "common support" requirement of the matching approach, 43% is quite large. This share is larger for matched girls.

ous,  $\Delta^{TS}$  slightly differs from zero for any twin pair. The results of the matching approach are unbiased if mean values of  $\Delta^{TS}$  are close to zero, which is the case here.

Immigrants and natives are further matched on values of the dummy variable "frequent participation". This could also bias results because matched immigrants and natives could differ in their levels of self-reported oral participation and their reporting behavior. For example, consider a pair of matched pupils whose "frequent participation" has value 1. It is possible that the level of oral participation is "high" for the immigrant and "very high" for the native. Possible differences in self-reported oral participation are therefore captured by the variable  $\Delta^{OP}$ , where "OP" stands for "oral participation". Like mean values of  $\Delta^{TS}$ , mean values of  $\Delta^{OP}$  are close to zero, too.

To investigate differences in the reporting behavior, one could calculate correlation coefficients between math test scores and oral participation. Since math skills are obtained from standardized tests, one could relate this "objective" measure to a self-reported variable. If correlation coefficients are computed separately for pupils from the treatment and control groups, similar correlation coefficients might indicate similar reporting behavior. In primary school, both groups of pupils behave similarly because  $\rho = 0.24$  for second-generation immigrants and  $\rho = 0.26$  for natives. Results in Figure 2 are therefore unlikely to be biased because of systematic differences in test scores, oral participation, or reporting behavior. For reasons of clarity, scatter plots are omitted in the remaining figures.

The remaining point estimates in Table 2a indicate that second-generation immigrants do not suffer from grade disadvantages in secondary school. The matching approach confirms this finding for second-generation immigrants in upper- and lower-secondary school because corresponding smoothed regression lines are not significantly deviating from the zero-grade-gap line (figures not reported).<sup>16</sup> Unlike class fixed effects estimates, the matching approach detects grade disadvantages for averagely performing second-generation immigrants in mid-

---

<sup>16</sup> All missing figures are available on request.

dle-secondary school (Figure 3). This finding is not biased because mean differences in grading-relevant characteristics are close to zero and correlation coefficients are similar ( $\rho = 0.15$  for second-generation immigrants and  $\rho = 0.14$  for natives).

Regarding girls, class fixed effects estimates in Table 2b suggest that girls in primary, middle-secondary, and lower-secondary school are not graded differently from boys. This is also confirmed by the matching approach (figures not reported). The only exceptions are girls in upper-secondary school who obtain better grades than comparable boys. Matching estimates in Figure 4 additionally reveal that all girls in upper-secondary school have grade advantages, regardless of their math achievement level.

## **5. The Difference between Detection and Explanation of Grade Discrimination**

Unlike math grades, math skills (as measured by test scores) are determined by other factors like preschool attendance or parental education.<sup>17</sup> Once differences in math skills and oral participation are accounted for, other pupil characteristics should not influence grading in math because this would not be in accord with the guidelines of the "Standing Conference of the Ministers of Education and Cultural Affairs of the Laender". If point estimates for second-generation immigrants or girls are affected by the inclusion of additional variables, these variables should therefore be interpreted as possible explanations for observed grade discrimination patterns.

Table 3a and b present class fixed effects estimates for various sets of control variables. Each cell contains a point estimate for second-generation immigrants or girls, respectively. Regardless of the used set of controls, the already described patterns remain: second-generation immigrants are graded worse in primary education and girls obtain better grades in upper-secondary school.

---

<sup>17</sup> See, for instance, Ammermueller (2007) and Woessmann (2004).

The first set of control variables in Table 3a and b are math test scores and oral participation (row I). These estimates are therefore identical with the results in Table 2a and b. The second rows in Table 3a and b also account for age and preschool attendance. Among these two additional controls, the estimate for pupil age is negative and significant in all school types (not reported in Table 3a and b). The third rows further control for the highest educational background of the parents.<sup>18</sup> In Table 3a, column 1, the p-value of the F-test for parental education is smaller than 0.01.

Compared to row I, point estimates for second-generation immigrants in primary school decrease by 7% in rows II and III. For girls in upper-secondary school, point estimates drop by 9%. Provided that all grading-relevant characteristics are sufficiently captured in row I, these additional variables (rows II and III) explain observed grade discrimination patterns to a small extent.<sup>19</sup>

Differences in reading skills are additionally controlled for in the fourth rows of Table 3a and b. Compared to row III, this is accompanied by a large drop (roughly 25%) in the point estimate for second-generation immigrants in primary education. The estimate for girls in upper-secondary school slightly decreases in row IV, too.<sup>20</sup> In Table 3a and b, row IV, estimates for reading skills are close to zero among secondary school tracks which implies that reading skills are not strongly correlated with math grades once math skills are taken into account.<sup>21</sup> Regarding girls in secondary education, the slight decrease of their point estimate is

---

<sup>18</sup> The categorical variable "parental education" captures the highest educational degree of the parents. It is classified by the "International Standard Classification of Education" (ISCED) system and ranges from 1 (ISCED 1) to 5 (ISCED 5). In regression analyses, it is decomposed into 5 indicator variables.

<sup>19</sup> In primary school, second-generation immigrants are 10.51 years old on average (s.d. = 0.47), whereas natives are aged 10.46 years (s.d. = 0.41). Girls in upper-secondary school are slightly younger than boys (15.44 years vs. 15.47 years). On average, second-generation immigrants enrolled in primary education have a lower socioeconomic background.

<sup>20</sup> Like math skills, reading test scores have mean 0 and standard deviation 1 within school types. If math grades in primary school are regressed solely on immigrant background, math skills, oral participation, and reading skills, the point estimate for second-generation immigrants becomes -0.186 (s.e. = 0.06) which is almost identical to the estimate in Table 3a, column 1, row IV. By contrast, regressing math grades (in upper-secondary school) on an indicator variable for female sex, math skills, oral participation, and reading skills, the point estimate for girls becomes 0.201 (s.e. = 0.04) which implies a small drop of 5% (compared to row I).

<sup>21</sup> In secondary school, point estimates for reading skills range from 0.02 to 0.05, and are insignificant in half of the cases. Significant estimates for reading skills are likely to be caused by the positive correlation between math and reading skills ( $\rho \approx 0.50$ ). At the same time, point estimates for math skills are between 0.34 and 0.41 (see Table 2a or b, columns 2-4). Reading skills are therefore grading-irrelevant in secondary school.

likely to be caused by the positive correlation ( $\rho = 0.12$ ) between female sex and reading skills.

Although second-generation immigrant background and reading skills are correlated ( $\rho = -0.22$  in primary school), too, collinearity is not the sole explanation for the sizeable change of the point estimate in the first column of Table 3a, row IV. Unlike secondary school, estimates for reading skills are large in primary education and highly significant.<sup>22</sup> To decide whether grading-relevant characteristics are additionally captured by reading skills in primary school, one must explain why reading skills matter in primary education but are irrelevant in secondary education at the same time.

There are at least two explanations for this pattern: (i) math skills could be insufficiently measured in PIRLS because the math test took 20 minutes (Bos et al. (2003): 196). In PISA, pupils were tested for a total of 210 minutes (Prenzel et al. (2004): 29). On the other hand, 20 minutes should be enough to evaluate basic math skills of fourth graders. Further, math tests in PIRLS are mainly in accordance with the German curriculum (Bos et al. (2003): 193). Therefore it is not clear whether grading-relevant characteristics are additionally picked up by the reading test in primary school. (ii) An alternative explanation for the "relevance" of reading skills in primary education considers institutional differences between the primary and secondary schooling levels: main subjects in primary school, namely math and German, are usually taught by the same teacher. Subject teachers are common in secondary school. If teachers in primary education do not clearly distinguish between a student's performance in math and German, math grades could be correlated with skills that are (solely) relevant in German. This reasoning would be consistent with the finding that reading skills have no impact on math grades in secondary education once math skills are controlled for. If this is the case, reading skills may not be considered as grading-relevant in math.

---

<sup>22</sup> In Table 3a, column 1, row IV, the point estimate for reading skills is 0.229 (s.e. = 0.03). Compared to row I, the estimate for math skills drops from 0.448 (s.e. = 0.02) to 0.331 (s.e. = 0.03).

## 6. Concluding Remarks

This paper examines whether second-generation immigrants and girls are graded worse than comparable natives and boys, respectively, once differences in grading-relevant characteristics are accounted for. To detect (positive or negative) grade discrimination, a matching approach from the labor market literature is implemented. Compared to class fixed effects estimates, matching provides more reliable results because pupils from the treatment and control groups are only taken into account if they attend the same class, have similar test scores, and similar levels of oral participation. Unlike OLS, matching does not assume a linear relationship between grades and test scores, which is a strong assumption.

Results of the matching approach suggest that second-generation immigrants with average and low test scores are affected by grade discrimination in primary education. The corresponding point estimate from class fixed effects regressions is  $-0.27$ , which is substantial because math grades range from 0 (fail) to 4 (very good). Second-generation immigrants enrolled at the most common secondary school tracks are not graded differently from natives, except averagely performing second-generation immigrants in middle-secondary school who are slightly graded worse than comparable natives.

Native girls and boys obtain same grades in math. The only exceptions are girls in upper secondary school. Once differences in grading-relevant characteristics are accounted for, matching estimates suggest that girls in upper-secondary school obtain better grades than comparable boys. The estimated grade advantage is  $0.211$ .

In Germany, track choice after primary education mainly depends on the average of grades in math and German. Hence grade disadvantages in primary education can bias track choice. The impact of these biases on future earnings are examined in Luedemann and Schwerdt (2010) for second-generation immigrants. As a final note, the matching approach indicates that *first*-generation immigrants, who are left out from the analysis, might also obtain worse grades in primary school. Since the confidence band of the corresponding



smoothed regression entirely overlaps zero-grade-gap line, this finding is not robust. Teachers and policy-makers should be aware of grade discrimination in primary education because immigrants account for roughly one third of the German student body.

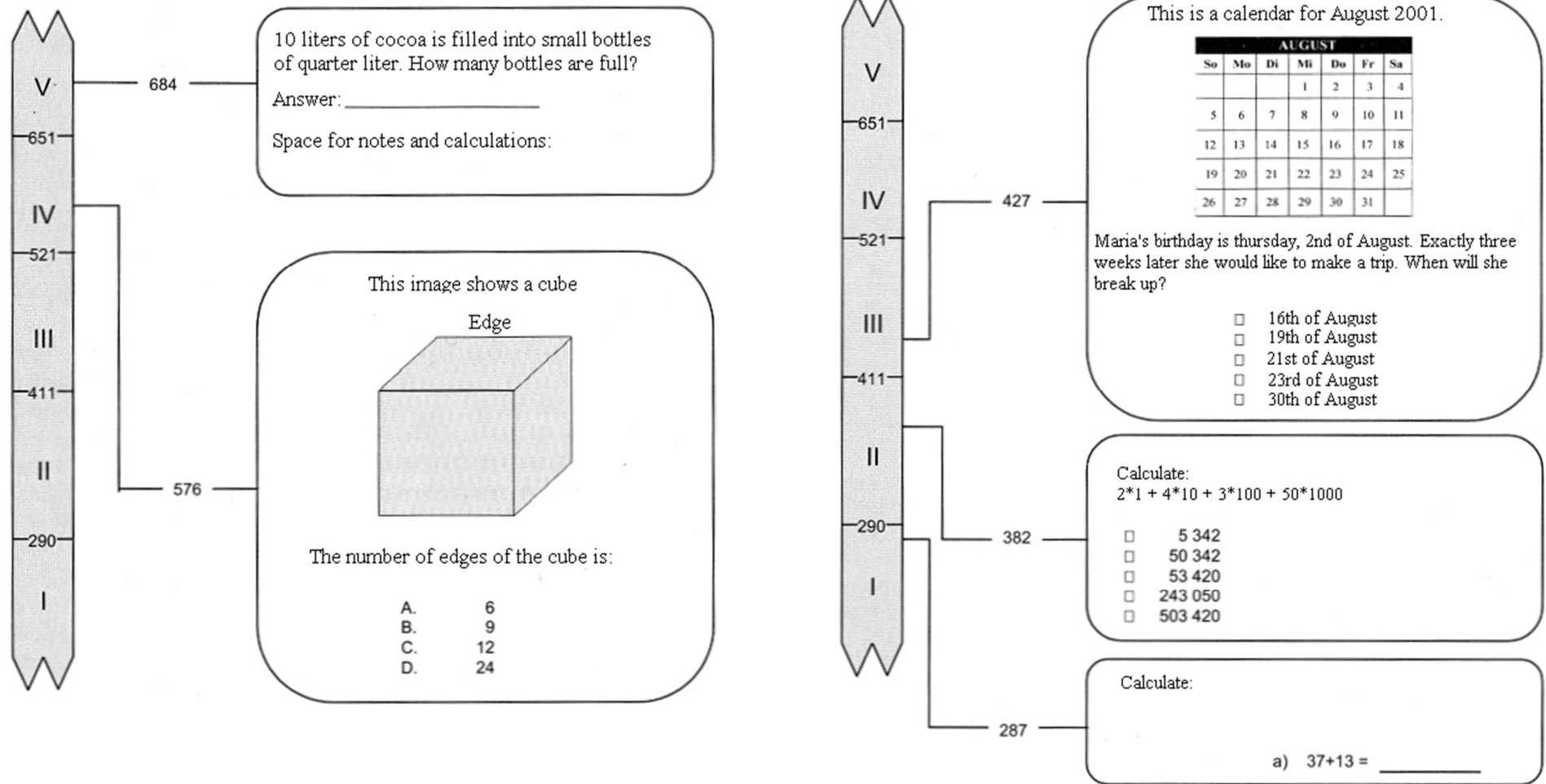
## References

- Altonji, Joseph G. and Charles R. Pierret.** 2001. "Employer Learning and Statistical Discrimination." *Quarterly Journal of Economics* 116(1), 313-50.
- Ammermueller, Andreas.** 2007. "Poor Background or Low Returns? Why Immigrant Students in Germany Perform So Poorly in the Programme for International Student Assessment." *Education Economics* 15(2), 215-30.
- Angrist, Joshua D. and Victor Lavy.** 1999. "Using Maimonides' Rule to Estimate the Effect of Class Size on Scholastic Achievement." *Quarterly Journal of Economics* 114(2), 533-75.
- Arrow, Kenneth J.** 1973. "Higher Education as a Filter." *Journal of Public Economics* 2(3), 193-216.
- Barsky, Robert et al.** 2002. "Accounting for the Black-White Wealth Gap: A Nonparametric Approach." *Journal of the American Statistical Association* 97(459), 663-73.
- Blinder, Alan S.** 1973. "Wage Discrimination: Reduced Form and Structural Estimates." *Journal of Human Resources* 8(4), 436-55.
- Bos, Wilfried et al.** 2003. "Erste Ergebnisse aus Iglu. Schuelerleistungen am Ende der Vierten Jahrgangsstufe im Internationalen Vergleich." Muenster.
- Dardanoni, Valentino, Salvatore Modica and Aline Pennsi.** 2009. "Grading across Schools." *B.E. Journal of Economic Analysis and Policy: Topics in Economic Analysis and Policy* 9(1), Article 16.
- Dee, Thomas S.** 2005. "A Teacher Like Me: Does Race, Ethnicity, or Gender Matter?" *American Economic Review* 95(2), 158-65.
- Dustmann, Christian.** 2004. "Parental Background, Secondary School Track Choice, and Wages." *Oxford Economic Papers* 56(2), 209-30.
- Himmler, Oliver and Robert Schwager.** 2007. "Double Standards in Educational Standards: Are Disadvantaged Students Being Graded More Leniently?" *ZEW Discussion Paper* 07(016), Mannheim: Zentrum für Europaeische Wirtschaftsforschung.
- Hinnerich, Bjoern Tyrefors, Erik Hoeglin and Magnus Johannesson.** 2010. "Are Boys Discriminated in Swedish High Schools?" *IFAU Working Paper* 2010(14), Uppsala: Institute for Labour Market Policy Evaluation.
- Hinnerich, Bjoern Tyrefors, Erik Hoeglin and Magnus Johannesson.** 2011. "Ethnic Discrimination in High School Grading: Evidence from a Field Experiment." *SSE/EFI Working Paper* 733, Stockholm: Stockholm School of Economics.
- Iacus, Stefano and Giuseppe Porro.** 2007. "School Grades and Students' Achievement: How to Identify Grading Standards and Measure Their Effects." *UNIMI Research Paper* 1061, Milan: Università degli Studi di Milano.
- Kultusministerkonferenz.** 2009. "Das Bildungswesen in der Bundesrepublik Deutschland 2008. Darstellung der Kompetenzen, Strukturen und bildungspolitischen Entwicklungen für den Informationsaustausch in Europa." Bonn.
- Lavy, Victor.** 2008. "Do Gender Stereotypes Reduce Girls' or Boys' Human Capital Outcomes? Evidence from a Natural Experiment." *Journal of Public Economics* 92(10-11), 2083-105.
- Luedemann, Elke and Guido Schwerdt.** 2010. "Migration Background and Educational Tracking: Is There a Double Disadvantage for Second-Generation Immigrants?" *CESifo Working Paper* 3256, Munich: Center for Economic Studies.
- Nopo, Hugo.** 2008. "Matching as a Tool to Decompose Wage Gaps." *Review of Economics and Statistics* 90(2), 290-99.
- Prenzel, Manfred et al.** 2004. "Pisa 2003. Der Bildungsstand der Jugendlichen in Deutschland - Ergebnisse des Zweiten Internationalen Vergleichs." Muenster.
- Riphahn, Regina T.** 2001. "Dissimilation? The Educational Attainment of Second Generation Immigrants." *C.E.P.R. Discussion Paper* 2903, London: Centre for Economic Policy Research.
- Schnepf, Sylke Viola.** 2007. "Immigrants' Educational Disadvantage: An Examination across Ten Countries and Three Surveys." *Journal of Population Economics* 20(3), 527-45.
- Spence, Michael.** 1973. "Job Market Signaling." *The Quarterly Journal of Economics* 87(3), 355-74.

- Sprietsma, Maresa.** 2009. "Discrimination in Grading? Experimental Evidence from Primary School." *ZEW Discussion Paper* 09(74), Mannheim: Center for European Economic Research (ZEW).
- van Ewijk, Reyn.** 2010. "Same Work, Lower Grade? Student Ethnicity and Teachers' Subjective Assessments." *Tinbergen Institute Discussion Paper* 127(3), Amsterdam: Tinbergen Institute.
- Woessmann, Ludger.** 2004. "How Equal Are Educational Opportunities? Family Background and Student Achievement in Europe and the United States." *CESifo Working Paper* 1162, Munich: Center for Economic Studies.
- Woessmann, Ludger and Martin West.** 2006. "Class-Size Effects in School Systems around the World: Evidence from between-Grade Variation in Timss." *European Economic Review* 50(3), 695-736.

## Tables and Figures

Figure 1a: Example exercises from PIRLS 2001

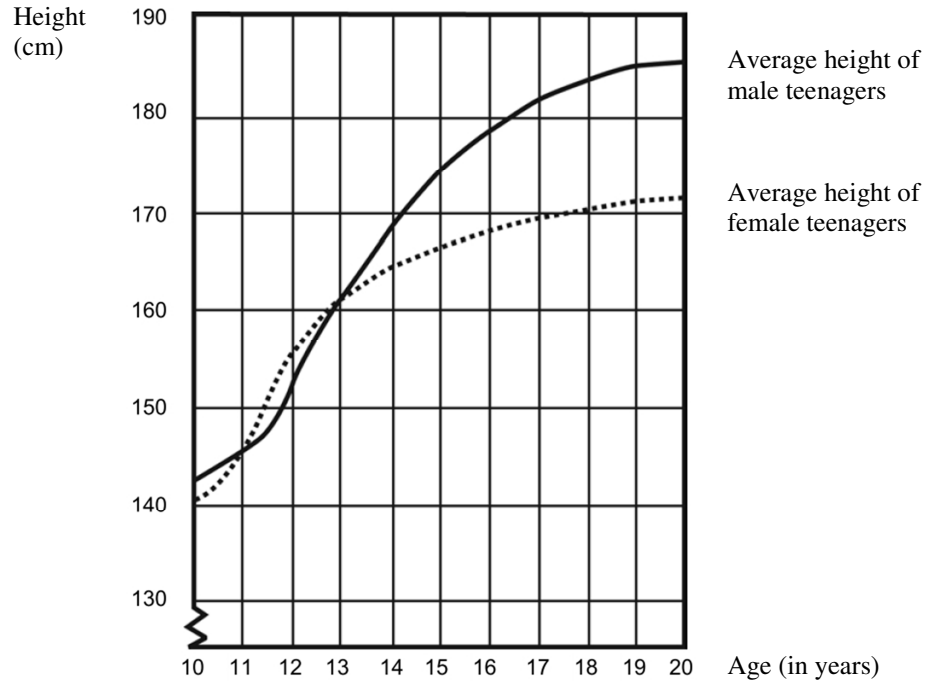


Source: Bos et al. (2003): 200. Correctly answered exercises are associated with skill levels as measured by the scale on the left side.

Table 1b: Example exercise from PISA 2003

**Teenagers become taller**

For the year 1998, the average height of male and female adolescents in the Netherlands is shown in the following figure:



**Question 1:**

Since 1980, the average height of 20-year-old women increased by 2.3 cm to 107.6 cm. What was the average height average height of a 20-year-old woman in 1980?

..... cm

**Question 2:**

According to the graphs, in which period(s) are female teenagers taller than male teenagers?

.....  
 .....

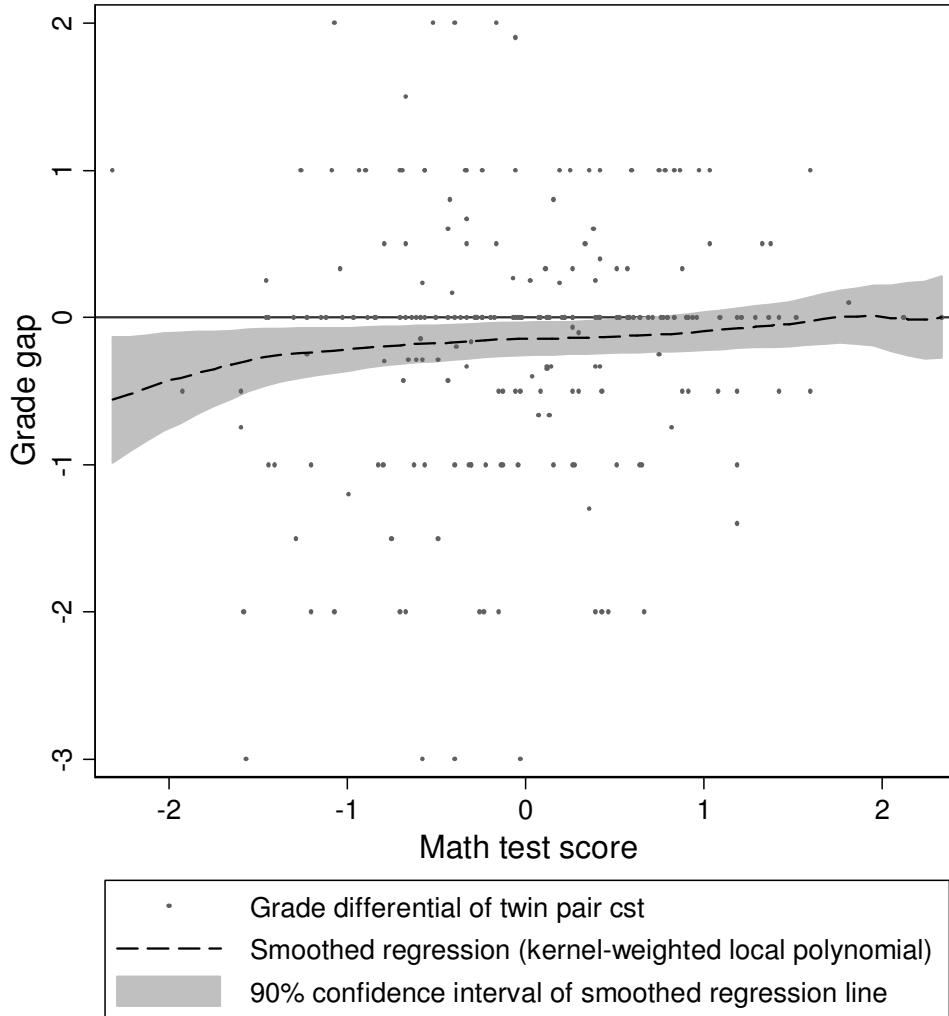
**Question 3:**

According to the graphs, which group of girls has smaller growth rates? Group A: "12-year-old girls or younger", or Group B: "13-year-old girls or older"? Explain your answer.

.....  
 .....

Source: [http://pisa.ipn.uni-kiel.de/Aufgaben\\_Mathe\\_neu3.pdf](http://pisa.ipn.uni-kiel.de/Aufgaben_Mathe_neu3.pdf).

Figure 2: Grade gaps of matched second-generation immigrants (SGI) in primary education

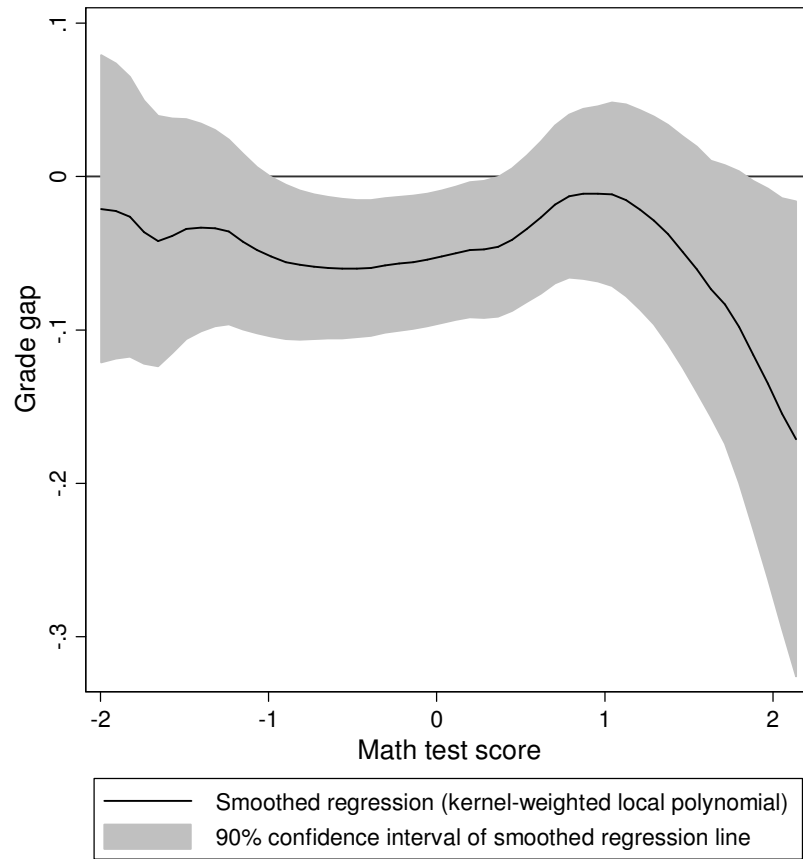


N(SGI): 152, N(statistical twin pairs): 219. Additional summary statistics for matched pupils:

Percentile:		0-33	34-66	67-100
		(1)	(2)	(3)
TS	mean	-0.884	-0.054	0.837
	s.d.	0.41	0.21	0.44
$\Delta^{TS}$	mean	-0.005	0.016	-0.019
	s.d.	0.17	0.15	0.15
$\Delta^{OP}$	mean	0.027	-0.075	-0.099
	s.d.	0.52	0.42	0.57

"SGI" means "second-generation immigrant". "TS" is the common test score,  $\Delta^{TS}$  is the test score difference, and  $\Delta^{OP}$  is the difference in self-reported oral participation of matched pupils. For any pair of matched pupils *cst*, the first pupil belongs to the treatment group and  $\Delta_{cst}^G \equiv G_{cst}^{treat} - G_{cst}^{cont}$ ,  $G \in \{TS, OP\}$ . The correlation between test scores and oral participation is 0.24 for (matched) second-generation immigrants and 0.26 for natives.

Figure 3: Grade gaps of matched SGI (middle-secondary school)

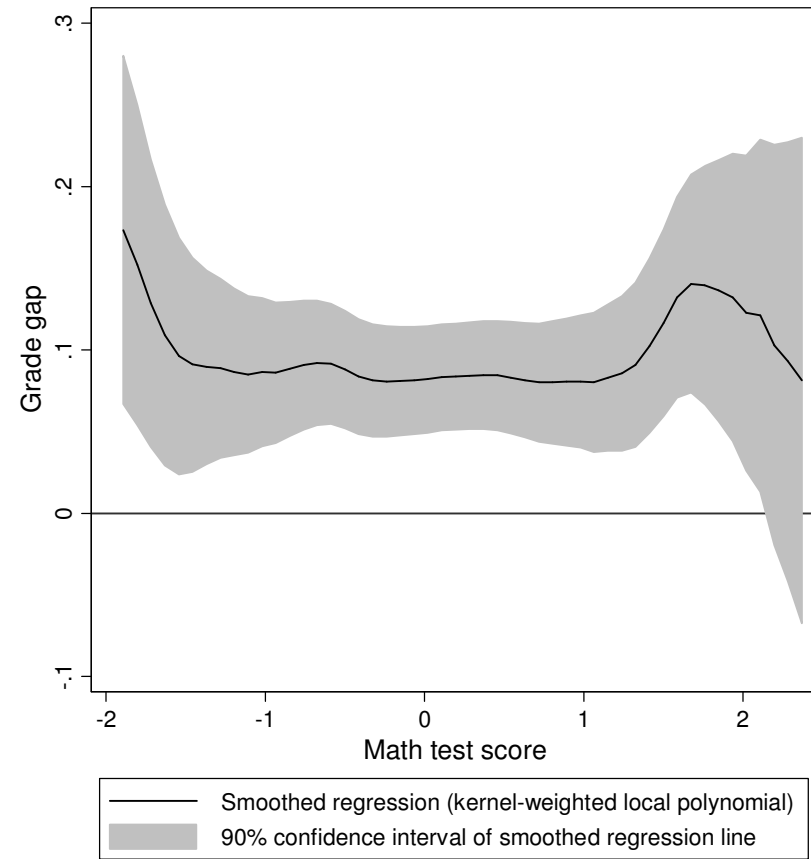


N(SGI): 202, N(statistical twin pairs): 560. Additional summary statistics:

Percentile:	0-33	34-66	67-100
$\Delta^{TS}$ mean	0.017	0.003	-0.031
s.d.	0.14	0.13	0.10
$\Delta^{OP}$ mean	-0.015	-0.004	0.008
s.d.	0.36	0.28	0.26

$\Delta^{TS}$  is the test score difference and  $\Delta^{OP}$  is the difference in self-reported oral participation of matched second-generation immigrants (SGI) and natives. The correlation between test scores and oral participation is 0.15 for SGI and 0.14 for natives.

Figure 4: Grade gaps of matched girls (upper-secondary school)



N(girls): 721, N(statistical twin pairs): 1427. Additional summary statistics:

Percentile:	0-33	34-66	67-100
$\Delta^{TS}$ mean	-0.006	-0.002	-0.014
s.d.	0.15	0.12	0.14
$\Delta^{OP}$ mean	0.050	0.034	-0.021
s.d.	0.44	0.35	0.34

$\Delta^{TS}$  is the test score difference and  $\Delta^{OP}$  is the difference in self-reported oral participation of matched girls and boys. The correlation between test scores and oral participation is 0.23 for girls and 0.28 for boys.

Table 1: Summary statistics (means and standard deviations)

School Type	Variable	Sample A: Natives and SGI			Sample B: Boys and Girls		
		(1) Natives	(2) SGI	(3) p-value	(4) Boys	(5) Girls	(6) p-value
Primary	Math Grade	2.53	2.11	0.00	2.60	2.47	0.00
	s.d.	0.86	0.95		0.82	0.90	
	Math TS	0.07	-0.36	0.00	0.14	-0.16	0.00
	s.d.	0.96	1.07		0.96	1.02	
	Oral Part.	3.26	3.11	0.00	3.34	3.19	0.00
	s.d.	0.76	0.8		0.78	0.74	
	N(pupils)	1347	354	1701	658	688	1346
	N(classes)	258	179	261	239	238	258
Upper-Sec.	Math Grade	2.02	1.92	0.35	2.02	2.02	0.92
	s.d.	0.99	1.11		1.00	0.98	
	Math TS	0.03	-0.27	0.00	0.26	-0.21	0.00
	s.d.	0.99	1.00		0.98	0.96	
	Oral Part.	2.23	2.13	0.30	2.32	2.16	0.00
	s.d.	0.89	0.92		0.94	0.84	
	N(pupils)	2312	213	2525	1055	1247	2302
	N(classes)	118	87	118	111	117	118
Middle-Sec.	Math Grade	1.91	1.81	0.13	2.00	1.83	0.00
	s.d.	0.99	1.03		1.01	0.96	
	Math TS	0.09	-0.45	0.00	0.26	-0.16	0.00
	s.d.	0.99	1.01		1.02	0.96	
	Oral Part.	2.37	2.42	0.32	2.50	2.27	0.00
	s.d.	0.95	0.99		0.99	0.89	
	N(pupils)	1745	257	2002	749	971	1720
	N(classes)	99	73	99	91	98	99
Lower-Sec.	Math Grade	2.00	1.86	0.04	2.10	1.86	0.00
	s.d.	1.01	0.98		1.05	0.92	
	Math TS	0.20	-0.45	0.00	0.21	-0.16	0.00
	s.d.	0.98	0.98		1.03	0.96	
	Oral Part.	2.44	2.56	0.24	2.58	2.25	0.00
	s.d.	1.04	1.05		1.03	1.03	
	N(pupils)	649	236	885	378	266	644
	N(classes)	80	67	81	78	73	80

Used data sets: PIRLS 2001 for primary school, PISA 2003 (subsamped by school track) for the secondary schooling level. Columns 1-3 contain information about the sample of natives and SGI (second-generation immigrants). Columns 4-6 report corresponding summary statistics for the sample of boys and girls. Columns 3 and 6 contain two types of information: The first three entries test for the difference in means between the control and treatment group. P-values are obtained from bivariate regressions of the variable of interest (i.e. math grade) on a constant and an indicator variable for immigrant background (or female sex). The remaining two entries in columns 3 and 6 are the total number of pupils and classes. The variable "math grade" ranges from 0 (fail) to 4 (very good). The higher a pupil's math test score, the higher his/her math proficiency. Math test scores have mean 0 and standard deviation 1 within school types. The categorical variable "oral participation" ranges from 1 (very low) to 4 (very high).

Table 2a: Class fixed effects estimates of immigrant background on math grades

Schooling level	Primary	Secondary		
Track	-	Upper	Middle	Lower
Second-generation immigrant (reference group: natives)	-0.271*** (0.07)	0.025 (0.07)	-0.034 (0.07)	0.043 (0.08)
Math test scores	0.448*** (0.02)	0.388*** (0.02)	0.343*** (0.03)	0.336*** (0.05)
Frequent participation	0.377*** (0.06)	0.568*** (0.04)	0.643*** (0.05)	0.600*** (0.06)
N(pupils)	1701	2525	2002	885
R <sup>2</sup> (adjusted)	0.298	0.256	0.239	0.214

Table 2b: Class fixed effects estimates of female sex on math grades

Schooling level	Primary	Secondary		
Track	-	Upper	Middle	Lower
Girl (reference group: native boys)	0.032 (0.05)	0.211*** (0.04)	0.084 (0.05)	-0.030 (0.09)
Math test scores	0.436*** (0.02)	0.402*** (0.02)	0.351*** (0.03)	0.406*** (0.05)
Frequent participation	0.346*** (0.06)	0.574*** (0.04)	0.610*** (0.05)	0.485*** (0.09)
N(pupils)	1346	2302	1720	644
R <sup>2</sup> (adjusted)	0.281	0.267	0.240	0.234

Dependent variable: math grade. Significance levels: \* 10%, \*\* 5%, \*\*\* 1%, depending on clustered (class level) standard errors, which are in parenthesis. Used data sets: PIRLS 2001 for primary education, PISA 2003 (subsampling by school tracks) for the secondary schooling level. The dependent variable "math grade" ranges from 0 (fail) to 4 (very good). The higher a pupil's math test score, the higher his/her math proficiency. Math test scores have mean 0 and standard deviation 1 within school types. The dummy variable "frequent participation" has value 1 whenever a pupil's level of self-reported oral participation is 3 (high) or 4 (very high). All regressions include a constant (results not reported).



Table 3a: Class fixed effects estimates of immigrant background on math grades (various sets of control variables)

Schooling level	Primary	Secondary		
Track	-	Upper	Middle	Lower
<i>Used set of control variables</i>				
I: math test scores, oral participation	-0.271*** (0.07)	0.025 (0.07)	-0.034 (0.07)	0.043 (0.08)
II: I, age, preschool attendance	-0.268*** (0.06)	0.033 (0.07)	-0.042 (0.07)	0.037 (0.08)
III: II, parental education	-0.254*** (0.06)	0.023 (0.07)	-0.087 (0.08)	0.028 (0.09)
IV: III, reading test scores	-0.185*** (0.06)	0.025 (0.07)	-0.072 (0.08)	0.030 (0.09)
<i>R<sup>2</sup> (adjusted)</i>				
I:	0.298	0.256	0.239	0.214
II:	0.309	0.269	0.243	0.217
III:	0.319	0.270	0.245	0.215
IV:	0.363	0.279	0.247	0.218
N(pupils)	1701	2525	2002	885

Table 3b: Class fixed effects estimates of female sex on math grades (various sets of control variables)

Schooling level	Primary	Secondary		
Track	-	Upper	Middle	Lower
<i>Used set of control variables</i>				
I: math test scores, oral participation	0.032 (0.05)	0.211*** (0.04)	0.084 (0.05)	-0.030 (0.09)
II: I, age, preschool attendance	0.012 (0.05)	0.194*** (0.04)	0.068 (0.05)	-0.045 (0.09)
III: II, parental education	0.016 (0.05)	0.193*** (0.04)	0.069 (0.05)	-0.039 (0.09)
IV: III, reading test scores	-0.052 (0.05)	0.182*** (0.04)	0.044 (0.05)	-0.066 (0.09)
<i>R<sup>2</sup> (adjusted)</i>				
I:	0.281	0.267	0.240	0.234
II:	0.290	0.279	0.242	0.238
III:	0.299	0.278	0.242	0.239
IV:	0.342	0.278	0.243	0.240
N(pupils)	1346	2302	1720	644

Dependent variable: math grade. Significance levels: \* 10%, \*\* 5%, \*\*\* 1%, depending on clustered (class level) standard errors, which are in parenthesis. Used data sets: PIRLS 2001 for primary education, PISA 2003 (subsamped by school tracks) for the secondary schooling level. Each cell contains a point estimate for second-generation immigrants and girls, respectively. The indicator variable "preschool attendance" has value 1 if a pupil attended preschool for more than one year. The categorical variable "parental education" captures the highest educational degree of the parents. It ranges from 1 (ISCED 1) to 5 (ISCED 5). It is decomposed into 5 indicator variables. Reading test scores have mean 0 and standard deviation 1 within school types. Apart from the various sets of controls in rows I-IV, all regressions include a constant and indicator variables for missing values in additional controls (results not reported).