Immigrant Concentration at School and Natives' Achievement: Does Length of Stay in the Host Country Matter?

Laurent Bossavie, *European University Institute (EUI)* *

Abstract

Using a rich dataset of primary school students in the Netherlands, this paper investigates the effects of immigrant concentration in the classroom on the academic achievement of natives. It exploits rare information on age-at-migration of school peers to estimate separate spillover effects by duration of stay of immigrant classmates. To identify treatment effects, it uses attractive features of the Dutch primary school system and cohort-by-cohort deviations in immigrant concentration within schools. While we find that the concentration of children who recently migrated negatively effects natives' performance, we report no impact of foreign-born students who have been in the country for a longer period. In addition, the adverse effects of recent migrant concentration in the classroom are primarily observed among disadvantaged natives. The importance of taking into account heterogeneity in the duration of stay of immigrant peers could explain the mixed findings reported by previous work.

1 Introduction

Given the sharp increase in international labor mobility and the recent rise in refugee inflows, national economies are facing the issue of economic integration of migrants to an unprecedented degree. While the economic consequences of immigration on the labor market have been widely

*Current address: The World Bank, 1818 H Street NW, 20433 Washington DC, USA. E-mail: lbossavie@worldbank.org. Tel: 202-751-6478.

studied, immigration may also affect schooling outcomes and human capital acquisition by natives. A growing body of literature, initiated by the seminal contribution of Lazear (2001), shows that classroom composition can impact individual school performance. Policy measures taken by some governments also suggest that the growing concentration of immigrant students in the classroom is of concern among policy makers. In 2010, the Italian Ministry of Education introduced a law that caps at thirty percent the share of foreign-born students in public school classrooms. Such measures, however, are largely motivated by anecdotal evidence of disruption rather than rigorous estimations. In addition, economic theory is inconclusive about whether immigrant concentration in the classroom produces positive or negative effects, if any, on the performance of natives. While it is plausible that a diverse student body has positive effects due to complementarities in abilities and types, a very heterogeneous class also makes teaching as well as peer interactions harder.[1]

Evidence on the impact of migration on the school system and human capital acquisition has been growing in recent years, but remains thin and reports mixed findings. Part of the literature finds no impact of immigrant concentration in the classroom on natives' achievement, while a comparable number of contributions report negative effects. At least three factors could explain these mixed results. First, variation in local contexts and in the capacity of local school systems to absorb immigrant children may play a role. Second, difficulties in identifying treatment effects can lead to either underestimate or overestimate spillovers by immigrant students. Third, different types of immigrant children may generate different spillovers on natives, and among natives, some categories of students might be more affected than others by the presence of immigrant classmates.

One important limitation of previous literature is that it typically treats immigrant children as

[1] See Lazear (2001) for theoretical insights on the topic and Duflo et al. (2011), among others, for an empirical application.

an homogeneous group. In particular, it does not take into account the duration of stay of foreign-born children in the host country when estimating peer effects. There are reasons to suspect that immigrant classmates who recently arrived to the host country generate different spillovers, if any, compared to children who have lived in the host country for a longer period. Duration of stay in the host country can proxy for several factors associated with immigrant students' achievement, as well as spillovers on their native classmates. Immigrants that recently arrived to the host country may have a weaker command of the local language, face initial difficulties associated with cultural assimilation, or experience emotional distress associated with recently moving to a new country. They may therefore require greater attention from teachers compared to immigrant children that arrived to the country at an earlier age. In that regard, contributions such as Ohinata and van Ours (2012) have shown a strong and positive association between duration of stay in the host country, and scholastic achievement of foreign-born students.

Given the limitations of previous literature, this paper contributes to the growing but still thin literature on the impact of immigrant peers on natives' scholastic achievement in several respects. First, it sheds light on the fact that the effect of immigrant concentration in the classroom depends on the duration of stay of immigrant students in the host country. By exploiting rare information on the during of stay of immigrant classmates in the Netherlands, it separately estimates the impact of foreign-born peers that recently arrived to the Netherlands from those that arrived in the country at an earlier age. To the best of our knowledge, this is the first contribution that distinguishes between different durations of stay of foreign-born peers when estimating the effects of immigrant concentration.[2] As the peer effect literature suggests that weaker students might be more strongly

[2]While Ohinata and van Ours (2016) distinguish between first and second generation migrants, this paper examines heterogeneity in duration of stay among first-generation migrant peers.

affected by classroom composition, we also investigate the heterogeneous effects of immigrant concentration on different socio-economic categories of natives.

Second, the paper takes advantage of some features of the Dutch primary school system and of the PRIMA dataset to identify the effect of immigrant peers on natives' scholastic achievement. Estimates based on classroom-level peer composition reported in the literature are likely to suffer from non-random allocation of students between classrooms.[3] On the other hand, using grade-level peer composition is likely to underestimate peer effects, as most learning spillovers are likely to occur at the classroom level (see Carrell et al. (2009) or Brodaty (2010), among others). The Dutch primary school system presents an attractive feature to tackle those issues, as the large majority of Dutch primary schools only have one classroom per grade. Although we report our main results for the full sample, we assess the robustness of our estimates in the subsample of schools with a single classroom per grade. Our identification strategy relies on small changes in immigrant concentration across grades within the same school, controlling for school-specific trends in immigrant concentration across cohorts. We run several tests to assess the validity of our identification strategy, including balancing tests for selection on observables, but also placebo tests which suggest that our results are not driven by selection on unobservables.

Finally, this study adds to the thin literature that investigates the effects of immigrant concentration on natives' achievement at school in early ages, as our sample consists of primary school students from age five. The focus on early ages is relevant in the specific context of the question investigated as immigrant classmates, defined as foreign-born students, have spent less time in the host country at those ages than older students. One could therefore expect greater disparities with

[3]One recent exception is Ballatore et al. (2015) which attempt to account for the endogeneity of classroom formation to identify the effect of immigrant classmates.

native children in those ages and potentially stronger learning spillovers. Studying this question for young children is also important as the literature highlights the cumulative role played by the acquisition of basic skills such as reading and simple arithmetics in fostering further skills and shaping labor market outcomes.[4]

Our results suggest that the impact of immigrant concentration on natives' test scores is heterogeneous, both in the type of immigrants that are part of the treatment, but also in the type of natives that are affected. While immigrant classmates who have already been in the Netherlands for some years are found to have no impact on natives' achievement, we report a negative and significant impact of the concentration of migrants that have been in the country for a short period. The effect size is however small in magnitude, and statistically significant only for scholastic achievement in Dutch language. Furthermore, we report that native students from higher socio-economic backgrounds are not affected by the concentration of immigrant classmates in their classroom, even if those are recent migrants. On the other hand, we find adverse effects of the share of recent migrant classmates on the scholastic achievement of natives with low parental education.

The paper is organized as follows. Section 2 reviews related literature on the topic. Section 3 provides background information on immigration and primary education in the Netherlands. Section 4 presents our data. Section 5 describes our identification strategy and provides supporting evidence for its validity. Section 6 presents our main results while Section 7 performs placebo tests and sensitivity checks. Section 8 concludes.

[4]See Cunha and Heckman (2007), among others.

2 Related Literature

This paper first relates to the broader literature on peer effects at school. The hypothesis that the behavior and outcomes of students are affected by their peers is formalized in the seminal contribution of Lazear (2001). The classroom is viewed as a public good in which classroom disruption by some students produces negative externalities on the entire class. As students are heterogeneous in their propensity to disrupt the class, changes in classmates composition affect instruction and individual achievement. From an empirical point of view, a large body of literature using both experimental and non-experimental methods has attempted to estimate the effects of classroom composition on individual school performance.[5]

Evidence on the impact of immigrant classmates on scholastic achievement is more scarce. In the US, a related literature studies the effect of ethnic segregation on academic achievement. Using data from Texas public schools, Hoxby (2000) and Hanushek et al. (2004) use variation in ethnic composition of adjacent cohorts in a given school to identify the effect of ethnic composition on student outcomes. Both studies find that the test scores of African-american students are negatively affected by the share of African-american classmates, while those of white students are unaffected by the percentage of black classmates. Using quasi-experimental evidence from the Metropolitan Council for Education Opportunity (Metco) in Boston, Angrist and Lang (2004) exploit the fact that students from disadvantaged neighborhoods were transferred by Metco to receiving schools to identify the effect of the share of minority classmates. They find no significant impact of an increase in the share of minority peers on the achievement of white students in math, reading, and language scores for 3rd, 5th, and 7th graders.

[5]Epple and Romano (2011) or Brodaty (2010) provide a literature review of applied work estimating peer effects in the classroom.

Despite the importance of immigration issues for European countries, the literature on the effect of immigrant peers on natives' achievement is still thin and reports mixed findings. This question was studied in the European context by Jensen and Rasmussen (2011), Brunello and Rocco (2013), Ohinata and van Ours (2013, 2016), Geay et al. (2013), Ballatore et al. (2015), Schneeweis (2015), and Tornello (2016).[6] While Ohinata and van Ours (2013), Ohinata and van Ours (2016), Geay et al. (2013) and Schneeweis (2015) report no effect on natives, other studies find statistically significant negative impacts. To the best of our knowledge, none of these studies distinguish between different durations of stay of foreign-born classmates when estimating those effects.

Jensen and Rasmussen (2011) examine this issue in the Danish context using test score data from the Project for International Student Assessment (PISA) at age 15, combined with Danish administrative data on neighborhood composition. To address the non-random selection of immigrants between schools, they instrument the share of immigrants in the school by immigrant concentration within a larger geographical area. They report a negative effect of immigrant concentration on the school performance of natives in both mathematics and reading, although estimated effects are small in magnitude.

Brunello and Rocco (2013) rely on cross-country differences in immigrant concentration among 27 European countries to estimate the effect of immigrant students on natives' achievement. They use test scores at age 15 from the Program for International Student Assessment (PISA) from 2000 to 2009, and their identification strategy relies on variations in immigrant concentration over time within countries, by aggregating micro-level data to the country level. Their results show a small

[6]Outside Europe, Gould et al. (2009) have also investigated the long-term impact of immigrant concentration in the classroom on the matriculation rates of natives in Israel.

negative effect of immigrant concentration on the school performance of natives, but estimate precision suffers from the small sample size due to data aggregation.

Ohinata and van Ours (2013) use data from the 2001 and 2006 Progress in International Reading Literacy Study (PIRLS), and the 1995 and 2007 Trends in International Mathematics and Science Study (TIMMS) in the Netherlands. They use variation in immigrant concentration across classrooms within the same school to identify the effect of having immigrant classmates on natives' test scores, and find no significant impact. Also in the Dutch context, Ohinata and van Ours (2016) use the PRIMA dataset to look at the effect of immigrant concentration at different parts of the test score distribution of native children. Using quantile regressions, they find no evidence for negative peer effects of immigrant children in any part of the distribution, after accounting for selection of migrants across schools.

Geay et al. (2013) use data on students at the end of primary school in England from 2003 and 2009. They rely on the influx of Eastern European migrants to the UK after 2005 to instrument the effects of immigrant concentration. They find virtually no effect of immigrant concentration in the classroom on English native speakers. Ballatore et al. (2015) use classroom formation rules in Italy as an exogenous source of variation in the share of immigrant classmates, in a sample of Italian primary schools. They find an adverse effect of the concentration of immigrant students in the classroom on natives' test scores in both language and mathematics. Schneeweis (2015), using Austrian primary school data, uses cohort-by-cohort variation in immigrant concentration within the same school to identify the treatment effect. She reports adverse effects of the share of immigrant classmates on the achievement of migrant students, but finds no impact on natives.

This paper also relates to the thin literature in economics looking at the impact of age at migration on the educational attainment of foreign-born students. In the Dutch context, Ohinata and van

Ours (2012) investigate the effect of age of migration on test scores of immigrant children at age 9 or 10, using data from the 2007 Trends in International Mathematics and Science Study (TIMSS). They find that immigrant children that entered at age 5 or older have a much lower science test score than children who entered as babies, suggesting assimilation effects. Other studies, such as Cortes (2006), Bohlmark (2008) or van Ours and Veenman (2006) focus on the impact of age at migration on educational attainment of first-generation immigrants at older ages.

3 Background and Institutional Setting

3.1 Immigrants in the Netherlands

In 2011, the Netherlands were populated by a population of 1.77 million immigrants, representing around 11 percent of the country population. As in most European countries, the majority of immigrants residing in the Netherlands come from lower-income countries. In 2011, the main groups of non-western origin populating the country were Turks (21%), Surinamese (19%) Moroccans (17%) and Antilleans (7%). Between 40 and 50% of these groups are second-generation immigrants. Almost one third of the Dutch immigrant population originates from former colonies, mainly Indonesia, Surinam and the Dutch Antilles. These immigrants had mostly a good command of the Dutch language when they entered the country, and were comparatively well-educated within school systems modeled on the Netherlands. A second immigration wave, consisting mostly of Turkish and Moroccans, entered the Netherlands in the 1960s. This second immigration wave was largely driven by an increased demand for low-skilled labor. Turkish and Moroccan immigrants came first as workers, and later for family formation and reunification. As a result, the large

majority of Turkish and Moroccan immigrants populating the Netherlands are from families with low educational backgrounds compared to native Dutch. In addition to these traditional groups, the Netherlands also hosts smaller immigrant groups from Iraq, Afghanistan or Iran.

The immigrant population is unevenly distributed across and within areas in the Netherlands. Non-western immigrants are considerably over-represented in the four major cities in the West of the country: Amsterdam, Rotterdam, The Hague and Utrecht. Approximately 50 percent of Surinamese and Moroccan immigrants live in one of the four major cities. Among the four major cities, Amsterdam and Rotterdam have the highest share of non-western immigrants with about 35 percent. Non-western migrants are also unevenly distributed within cities. In some districts of Amsterdam, 75 percent or more of young people are from a non-Western origin, while relatively few immigrants reside in city centers.

The uneven distribution of immigrants across cities and neighborhoods is reflected in the primary school system. In Amsterdam for example, 127 of the 201 elementary schools have more than 50 percent of children with a migration background, and 102 schools have a concentration of more than 70 percent. In contrast, in the nine suburban municipalities within a short distance from one of the most segregated districts of Amsterdam, only one school hosts more than 50 percent of children of non-western parents with low parental education.

3.2 The Dutch primary school system

From age five, all children residing in the Netherlands are legally required to attend school. Dutch primary schooling consists of eight grades covering age groups from four to twelve. Contrary to most European countries, school choice is free in the Netherlands. Parents are not restricted to send

their children to a school in a particular district, and are legally entitled to choose a school for their children, regardless of the neighborhood they live in. The primary school system consists of both public-authority and private schools that are both funded by the state. Both types of school receive, on top of their regular budget and based on the overall number of students, additional funding from the Ministry of Education on the basis of the percentage of immigrant students in their school population. The amount of additional funding is based on the total sum of weights assigned to students from different socio-economic categories enrolled in the school. The majority of students, children of Dutch middle class parents, receive a weight of 1. Children of Dutch parents with low levels of education are allocated a weight of 1.25. Bargee's children are weighted 1.4 and children of itinerant parents 1.7. Finally, children of immigrant parents with low education receive the highest weight of 1.9. Schools have a great amount of freedom in deploying the extra staffing hours, for instance by reducing class size, offering remedial teaching or appointing classroom assistants. The additional funding can also be used to introduce more specific measures, such as school-wide language policies or reception facilities for newcomers.

4 Data and Descriptive Statistics

4.1 The PRIMA data set

We constructed our panel of primary schools from six successive waves of the PRIMA longitudinal survey in the Netherlands. The survey was carried out every two years from 1994 to 2004 to follow the development of cognitive and non-cognitive skills of students throughout primary school. Participating schools were chosen to be representative of the entire population of Dutch primary

schools.[7] As we have multiple observations per school, we pooled all grades and years to exploit within school variation in the proportion of immigrant students. We linked the successive waves of PRIMA to build a panel of Dutch primary schools, observed in grade two, four, six and eight every two years from 1994 to 2004. We obtain a panel of about 600 schools with 12,053 grade-level observations.[8]

The data collected in PRIMA is based on answers to detailed questionnaires filled by teachers, parents, and school principals. As a result, the dataset contains rich information at the student, classroom and school levels. In particular, it contains detailed information on students' socio-economic and migration background. It reports whether the student is foreign born, the length of stay in the Netherlands, as well as the country of origin of the parents. We categorize as immigrants individuals for which the answer to the question "How long has the child been living in the Netherlands" is not "always". Contrary to most work in the literature, our definition of immigrants is therefore restricted to first-generation migrants that were born abroad, and does not include second generation migrants.

Student performance is measured by tests administered by the Dutch National Institute for Educational Measurement in Dutch language and mathematics. These tests were developed by the Dutch government testing agency to measure students' readiness in the two topics. We standardize individual raw test scores in the dataset so that the mean is 50 and the standard deviation is 10. Within each classroom, all students are sampled as long as they are present the day of the test. Contrary to many educational datasets used for peer effect estimation, an attractive feature of the

[7]The full PRIMA dataset consists of a representative sample of about 420 schools and also includes an additional sample of about 180 schools with children from a low socio-economic background.

[8]We refer to a grade-level observation as a grade of a given school observed in a given year. For example, grade 2 of school 1 observed in 1994 is a grade-level observation.

PRIMA dataset is that very few values are missing for the variables of interest. This allows to significantly alleviate the issue of non-random missing values in classroom peer data outlined by Sojournier (2013).

4.2 Descriptive Statistics

Table 1 and Table 2 report student-level and grade-level summary statistics of our sample, respectively. Table 1 shows that immigrant students have lower parental education compared to native students, as it is the case in most European countries. More than 43 percent of immigrant children have a father that did not study beyond primary school, as opposed to only 15 percent of native Dutch students. The proportion of immigrant students whose father did not study beyond primary school is particularly high among Turkish and Moroccan immigrants, which account for around one fourth of the total number of immigrants in our sample. 67 percent of Moroccan and Turkish students have a father that did not study beyond primary school, while this proportion is only 29 percent for immigrants from other countries. Table 1 shows that immigrant children in the sample perform on average significantly worse than native Dutch students, both in arithmetic and Dutch language tests. In addition, the achievement gap between native and immigrant students remains once we condition for parental education. This gap shows at all levels of parental education, and is larger in the subsample of Moroccan and Turkish immigrants.

[Table 1 about here]

Table 2 reports student characteristics and outcomes aggregated at the grade level, by level of immigrant concentration. We refer to grade-level observation as the set of students in grade g of school s, in year y. We observe significant selection of native students between grades with dif-

ferent levels of immigrant concentration. As expected, natives from more disadvantaged families tend to concentrate in grades where the fraction of immigrant students is high. The share of native students with a father that did not study beyond primary school ranges from 11 percent in grades with no immigrant to more than 37 percent in grades with more than 50 percent of immigrant students. The academic achievement of natives is also lower in grades with a high fraction of immigrant students. On the other hand, there is no clear pattern regarding the average achievement of immigrant students in school cohorts with different immigrant concentrations.

[Table 2 about here]

5 Empirical Strategy

5.1 The Identification Problem

The seminal contributions of Manski (1993) or Sacerdote (2001) have evidenced the fundamental problem of selection into peer groups which can contaminate peer effect estimates. In our context, it is likely that students exposed to a higher treatment intensity, i.e. with a higher share of immigrant children in their classroom, are also more likely to come from families with lower socio-economic status. Those are likely to obtain lower scores in achievement tests compared to students who have less immigrant in their classroom even if the treatment intensity was the same, which poses a fundamental identification problem.

The most obvious component of selection occurs between schools. Schools draw students from different neighborhoods and family backgrounds, leading to a concentration of students with similar characteristics in the same school. It is therefore crucial to use within-school variation

to identify the causal effect of immigrant concentration in the classroom on the achievement of natives.

A second type of selection of native and immigrant students into classrooms occurs within schools. Once school-fixed effects are accounted for, estimation of the effect of immigrant concentration might still be inconsistent if the allocation of students to classrooms within the same school is not random. School directors, teachers, or parents may indeed allocate students to classrooms in a non-random fashion, according to student characteristics that may not be observed by the researcher. Contrary to selection between schools, this second type of selection has received little attention in the literature, and is also more difficult to address. One notable exception is Ballatore et al. (2015) who attempt to account for the endogeneity of classroom composition according to migrant status using rules of classroom formation in Italy.

Carrell et al. (2009) also show that estimates for peer effects greatly differ depending on the accuracy with which econometricians identify the set of relevant peers. Estimating peer effects at the classroom level typically yields larger estimates, but one can doubt of the exogeneity of classroom formation outside the experimental setting. It seems natural, however, to expect that a significant fraction of peer effects in learning arises at the classroom level, since classes are the basic unit where learning takes place. Therefore, using a grade-level measure of immigrant concentration may generate a downward bias in the estimation of the treatment effect due to measurement error, as outlined by Brodaty (2010).

5.2 Identification of Immigrant peer effects

We are able to exploit one desirable feature of the Dutch context to tackle these issues. Dutch primary schools are on average of small size, and the large majority of schools only have one classroom per grade–level. In 2010, the average number of students enrolled by Dutch primary schools was 220 according to the Dutch Ministry of Education, which represents approximately 27.5 students per grade level. This figure is slightly lower in our sample of schools where the average number of students per grade is 26.3 (Table 2). In about 70 percent of the grade-level observations in our sample, students enrolled in the same grade are in the same classroom. While we conduct our baseline estimation on the full sample of schools, we also report our results for schools with a single classroom per grade, to assess the robustness of the estimates.

To address the potential endogeneity of students allocation to classrooms, we measure treatment intensity by the fraction of immigrant students in the grade, instead of using classroom-level peer measures. Our identification strategy therefore follows the spirit of Hoxby (2000), or Lavy and Schlosser (2011). We use the fact that several grades are observed within the same school, and rely on variation in immigrant concentration across cohorts to identify the effect of the share of immigrant peers on the learning outcomes of natives in the same grade. In other words, we examine whether the outcomes of native students across grades within the same school and year change systematically with the proportion of immigrant students in the same grade.

The inclusion of school fixed effects accounts for the most obvious source of student sorting between schools. This selection is likely to be particularly acute in the Netherlands, where a free school choice policy applies. In addition, there might also be some school-specific time varying factors that affects both students' outcomes and immigrant concentration. For example, school

administration might change from one year to another and affect both immigrant concentration as well as test scores. In addition, as outlined in section 3.2, primary school budget in the Dutch context is directly tied to the school socio-economic composition, which can vary across years. As school resources are likely to affect students' outcomes and are directly determined by the share of migrant in a school in a given year, it is crucial to control for year-specific school effects. We therefore add a a full set of school-year fixed effects γ_{sy} to our specification.

Since the test scores of students in the same grade are likely to be correlated and may therefore deflate standards errors, we follow the approach of Angrist and Lavy (1999) by using grade-level aggregates for estimation instead of individual-level data. We collapse individual observations to grade-level averages and estimate the effect of the share of immigrants in the grade on the average test score of native students. Using our panel of schools observed in four different grades over several years, we estimate the following reduced-form equation:

$$\overline{Y}_{sgy} = \alpha_g + \gamma_{sy} + \beta I_{sgy} + \rho \overline{X}_{sgy} + \varepsilon_{sgy}, \tag{1}$$

where s denotes the school, y denotes the year, and g the grade. $\overline{Y}_{sgy}$ denotes the average test score of native students in grade g of school s in year y. α_g is a grade effect, and γ_{sy} is a school-by-year effect. $\overline{X}_{sgy}$ is a vector of grade characteristics that is not necessary for the estimation if grade-by-grade changes in immigrant concentration within the same school year are exogenous, but it is added to the specification as a robustness check. I_{sgy} is the proportion of immigrant students in grade g of school s in year y. We are interested in consistently estimating β, which captures the effect of immigrant concentration in the grade on the average test score of native students.

Even after controlling for school-by-year fixed effects, one might still be concerned that vari-

ation in immigrant concentration across grades within schools is correlated with unobservable time-varying factors. Students in different grades within the same school started primary school in different years and it is therefore possible that changes in immigrant concentration across grades within schools also reflect endogenous changes in neighborhood population, or students' mobility. To alleviate this concern, we follow Hoxby (2000) and add to our baseline equation a full set of school-specific linear trends. Our reduced-form equation to estimate the effect of immigrant concentration in the grade therefore becomes:

$$\overline{Y}_{sgy} = \alpha_g + \gamma_{sy} + \sigma_{sy} grade + \beta I_{sgy} + \rho \overline{X}_{sgy} + \varepsilon_{sgy} \tag{2}$$

β is therefore identified from the deviations in the proportion of immigrant students in the grade from its linear trend across grades within the same school. The identifying assumption is that, once we allow for linear trends in immigrant concentration across grades, remaining changes in the share of immigrant students by grade are driven by factors that are exogenous to natives' test scores, such as the distribution of immigrants' birth year in the neighborhood. In other words, while the proportion of immigrant students in a school is relatively stable over time, there exists cohort-by-cohort variations that are purely driven by exogenous factors.

One potential threat to the identification strategy is the fact that families might react to changes in immigration concentration within the same school by moving away their children from the school. However, while parents may know the average immigrant composition of a given school, it is very difficult to predict the exact composition of a particular grade. In particular, the exact fraction of immigrant students enrolled in a particular school grade is unknown to parents before the beginning of the school year, and school departures are typically not allowed once the school

year has already started. In that regard, our identification strategy uses significantly more information than parents typically have to identify variations in immigrant concentration across grades within the same school.

Another potential threat to identification is grade retention which might affect immigrant and native students at different rates, and therefore potentially lead to non-random variation of native students' characteristics across grades that have different concentrations of immigrants students. In the next section and in section 7, we provide evidence suggesting that this is not the case, and that our key results are not driven by grade retention or students' selection based on observables and unobservables.

5.3 Balancing Tests

To investigate potential non-random variation in immigrant concentration across grades, we regressed our treatment variable, i.e. the fraction of immigrant students, on the characteristics of native students in the same grade and other grade characteristics. Table 3 reports the results of these balancing tests, where the fraction of immigrants in the grade is regressed on each of the measures of native students' socio-economic background and other grade characteristics, in separate regressions. Column 1 presents the results of a naïve benchmark OLS regression controlling for year and grade effects. The naïve estimates show a large and significant association between natives' observable characteristics, in particular parental education, and the percentage of immigrants in the grade. Correlations between immigrant concentration and natives' parental education are large in magnitude, and significant at the one percent level. As evidenced earlier, natives with low parental education tend to concentrate in schools with a high fraction of immigrant students.

Column 2 shows that the inclusion of school fixed effects reduces dramatically the magnitude of those correlations. All estimates become statistically insignificant, with the exception of natives whose parents achieved upper secondary education which remains marginally significant, as well as the total number of students enrolled in the grade. Using within-school variation in immigrant concentration therefore significantly alleviates selection issues. Once school fixed effects are accounted for, there is little remaining association between immigrant concentration and grade characteristics.

Column 3 shows the association between the share of immigrants in the grade and natives' characteristics when school-by-year fixed effects are controlled for. This specification further controls for school-specific year effects to account for idiosyncratic shocks that could affect a school in a given year and may be correlated with immigrant concentration, as well as year-specific school financial resources. Controlling for school-specific year effects further decreases the magnitude of the correlations, which become virtually zero and insignificant for all grade-level characteristics included in the test.

Finally, Column 4 shows the association between grade characteristics and the fraction of immigrants resulting from our identification strategy, controlling for school linear time trends in immigrant concentration. The magnitude of all correlations are virtually zero and very similar to the school-by-year fixed effect estimates, but the magnitude of the correlation with enrollment is further reduced. This suggests that the variation in immigrant concentration resulting from our identification strategy is uncorrelated with changes in observables relevant for achievement. We repeated this exercise for the share of recent immigrants, defined as foreign-born students that have been in the Netherlands for less than four years. The results are reported in Table A1 and also show that the association between the share of recent immigrants in the grade and other observable

grade-level characteristics is virtually zero.

Our identification strategy requires the fraction of immigrants in the grade to be uncorrelated to both observable and unobservable grade characteristics. As emphasized by Gould et al. (2009), this type of balancing test does not provide a proof for random assignment. However, the lack of association between treatment and other correlates of academic achievement resulting from our identification strategy suggests that unobservables are also unlikely to be correlated with treatment intensity, especially if those unobservables are correlated with observables. Overall, the sharp contrast between the naïve estimates and those resulting from our identification strategy shows the extent to which it eliminates the bias stemming from selection. To further alleviate concerns of remaining spurious correlations between immigrant concentration in the grade and unobservables, we also conduct in section 7 placebo treatment tests and additional robustness checks suggesting that this is not the case.

[Table 3 about here]

6 Results

6.1 Effects of Immigrant Concentration

Row 1 of Table 4 report the linear effects of the share of immigrants in the grade on the average test score of natives (Treatment 1). This is the standard treatment effect typically estimated by the literature. According to the baseline estimates, immigrant concentration in the grade has a negative impact on natives' test scores in language and mathematics. These negative effects are however statistically insignificant, even in a context where grade-level peer estimates are unlikely to lead

to classical measurement error as the large majority of schools in he Netherlands only have one classroom per grade. The estimated effect size is low in magnitude: an increase by 10 percentage points in the share of immigrant classmates in the grade reduces the average verbal test score of natives by less than 0.10, compared to a standard deviation of 5.4 in natives' average language test score. The estimated effect is even smaller for mathematics test scores. The inclusion of the full set of grade mean characteristics as controls has little impact on the effect size, as expected in a quasi-experimental setting. Therefore, when we treat immigrants peers as an homogeneous group, we find no impact of immigrant concentration on natives' achievement. Although we use a different source of variation in immigrant concentration, these findings are consistent with Ohinata and van Ours (2013) and Ohinata and van Ours (2016) in the Dutch context.

6.2 Effect of Immigrant Concentration by Duration of Stay

Existing evidence suggests that young immigrant children who have been in the country for longer tend to better perform better in school compared to immigrant children that have been in the country for a shorter period of time (Ohinata and van Ours (2012)). Recent arrival to the country may generate emotional distress associated with cultural adjustment, and may also require acquisition of the host country language. During this time, recent migrant students may require additional teaching resources, which could leave fewer resources for native children studying in the same classroom. This effect is likely to be less pronounced when immigrant children have already spent substantial time in the country, acquired a stronger command of the host country language, and started to assimilate to the local context. Therefore, the negative effect of the concentration of immigrant peers that recently migrated might be larger than that of foreign-born students that have

been in the host country for a longer.

[Table 4 about here]

To test this hypothesis, we exploit rare individual-level information on the length of stay of foreign-born students in the Netherlands available from the dataset. We classify as recent immigrants foreign-born children that have been in the country for less than four years, which is the median duration of stay of first-generation immigrants in our sample. We then estimate the effect of two alternative treatments: the share of recent immigrants in the grade (treatment 2), and the share of immigrants who have been in the country for a longer time (treatment 3). Row 2 and 3 of Table 4 report estimates for these two alternative treatment effects. They show that the share of recent immigrants in the grade has a negative and statistically significant effect on natives' verbal test scores. The estimated effect size remains relatively small in magnitude. According to our estimation, an increase of the share of recent immigrants by 10 percentage points reduces natives' average language test score by -0.30, about 0.06 standard deviation. The estimated effect on natives' outcomes in mathematics is also negative, but the effect size is smaller and statistically insignificant. Estimates for the effect of the share of long-term immigrants in the grade show virtually no effect of the treatment on natives' test scores in both language and mathematics.

6.3 Heterogeneous Effects by Natives' Types

We previously assumed that the effect of immigrant concentration was identical for all natives. However, the literature on classroom peer effects suggests that spillovers might be heterogeneous across student types. In particular, weak students are typically found to be more responsive to their peer composition than students from less disadvantaged backgrounds. Hanushek et al. (2003)

find that the performance of students in the lower end of the ability distribution is more negatively impacted by the presence of repeaters in their grade. To investigate this possibility in our context, we look at the impact of immigrant concentration on two types of natives. We investigate separately the impact on natives with low parental education and high parental education, as proxies for family background and socio-economic status.

We run the same regressions as in Table 4 separately for these two groups of natives. Results are presented in Table 5. Among natives with high parental education, estimated treatment effects are approximately -1 for mathematics and language, and statistically insignificant. Among native students with low parental education, estimated effects on language and mathematics test scores are both negative, and larger in magnitude compared to natives with high parental education. The estimated treatment effect is approximately 3.35 for Dutch language test scores, and significant at the 5% level. For mathematics, estimates are statistically insignificant. This indicates heterogeneity in treatment effects depending on the socio-economic background of native students receiving the treatment. While natives with high parental education are unaffected irrespective of the duration of stay of immigrant classmates, the scholastic achievement of natives with low parental education appears to be adversely impacted by the presence of recent immigrants.

[Table 5 about here]

7 Robustness Checks

7.1 Falsification Tests

To further check that our estimates do not capture a spurious correlation between immigrant concentration and other grade-specific factors, we conduct falsification tests with placebo regressions. Instead of regressing native students' outcomes on the true concentration of recent immigrants in the grade (actual treatment), we estimate regressions in which the treatment measure is replaced by the share of recent immigrants in the previous grade, or in the next grade (placebo treatments). If native students' outcomes are affected by grade-specific unobservables correlated with immigrant concentration at the school level, then the placebo should also be significantly associated with outcomes. Finding a significant effect of the placebo on test scores would therefore cast doubt on the validity of the identification strategy.

Results reported in Table 6 show no association between the share of immigrants in the previous or next grade and native students' test scores. Estimates of placebo effects are much smaller than for the actual treatment, statistically insignificant, and of inconsistent signs. For example, when using the presence of immigrants in the next grade (placebo 1) instead of the actual concentration of immigrants in the grade, the estimated effect on natives' language scores is -0.36 (standard error: 1.24), compared to -3.08 with the actual treatment. When the proportion of immigrants in the previous grade is used as alternative placebo (placebo 2), the estimated coefficient is of the opposite sign, and also statistically insignificant. This can be viewed as further evidence that our estimates capture the true effect of immigrant concentration on students' outcomes, rather than the confounding influence of grade-specific characteristics. In particular, if endogenous student mobility was driving our results, we would expect the share of immigrants in previous grades to be

a significant predictor of current achievement. The results of our placebo regressions suggest that this is not the case.

[Table 6 about here]

7.2 Using variation across years in the same grade within schools

One potential threat to identification in the Dutch context is the non-random allocation of immigrant students across grades within the same school through grade retention. Grade retention is a relatively common phenomenon in the Netherlands where students performing weakly can be encouraged to repeat a grade. Repetition rates are likely to differ between native and immigrant students and can therefore lead to non-random allocation of immigrant and native students across grades within the same school. If a school tends to hold immigrant students back so that more of them are placed in the grade with better or worse native students compared to the adjacent grade, then our previous results could be biased.

We showed in Table 3 that the association between the share immigrants and the share of repeaters in the grade is virtually zero once school-by-year effects are accounted for. Table 6 also alleviated concerns about selection of immigrant students based on unobservables across grades within the same school. To further alleviate concerns, we check the robustness of our key results to using only one grade per school, by exploiting variation across years in the same school to identify the treatment effect.[9] We also control for a linear time trend in immigrant concentration within schools, to account for a potential school-specific trends in immigrant concentration.

Results are displayed in Table 7. Our key findings are robust to restricting the sample to one

[9]We report the results for grade 4. Estimates for grade 6 and 8 are similar and reported in table A2 of the Appendix.

grade and exploiting variation in immigrant students concentration across years within the same school. Using this alternative identification strategy, the share of recent immigrant in the classroom negatively affect natives' achievement in language, while foreign-born students who have been in the country for longer are found to have no impact on natives. As in our baseline estimates, we do not find any effect on mathematics test scores. The effect size on language scores is also very similar to the one estimated with our preferred estimation strategy.

[Table 7 about here]

7.3 Restricting the Sample to Schools with One Classroom per Grade

Our baseline estimates use grade-level peer composition to identify the causal effect of immigrant students in the classroom on the achievement of natives. As detailed earlier, the potential bias associated with using grade-level measures as opposed to classroom-level measures is greatly attenuated in our context as most primary schools in the Netherlands only have one classroom per grade. We however assess the robustness of our findings in the subsample of schools with a single classroom per grade, which represent approximately 70% of our sample of schools.

Estimated effects of the concentration of recent migrants in the two samples are displayed in Table 8. The estimated effect of the concentration of recent immigrants in the grade on natives' language test scores is negative and significant at the 1% level in both subsamples. The effect size is also very similar, although estimates are slightly larger in the restricted sample for language, and very similar for mathematics. The slightly smaller effect size in language could result from a residual downward bias in the estimation of spillovers in schools that have more than one classroom per grade. Alternatively, it could also originate from migrant spillovers being actually larger in

smaller schools because, for example, they might be lacking adequate structures to accommodate recent migrants.

[Table 8 about here]

8 Conclusion

Our findings contribute to the literature on immigrant peer effects in the classroom by showing that the magnitude of spillover effects greatly depend on the duration of stay of first-generation immigrant classmates in the country. Our results in the Dutch context suggest that immigrant students that have been living in the country for a short period negatively impact natives' performance in language. In contrast, the share of foreign-born classmates who have already been living in the country for a longer period is found to have no effect. Ignoring this distinction when estimating immigrant peer effects can therefore be misleading, which could partly explain the mixed findings reported by the literature on the topic.

Although the exact mechanisms behind these results need to be further investigated, our findings indicate that assimilation and host country language acquisition may play a role in generating immigrant peer effects in the classroom. If heterogeneity among classmates drives learning spillovers as suggested by Lazear (2001), and if immigrant students progressively assimilate and acquire a greater command of the host language over time, it is plausible to observe learning spillovers decline with the duration of stay of immigrants in the host country. The fact that adverse effects are only statistically significant for language test scores also points towards host country language proficiency as a potential channel.

The adverse effects of recent migrants on natives that we identified are however relatively small

in magnitude. An increase by 10 percentage points in the share of recent migrants in the classroom is estimated to reduce natives' language test scores by about 0.06 standard deviation. The specificities of the Dutch primary school system and the features of our dataset provide comfort on the precision of our estimates. The predominance of schools with a single classroom per grade in the Netherlands allows to circumvent the issue of non-random allocation of students to classroom by using grade-level measures of peer composition, while alleviating concerns about attenuation biases resulting from measuring peer composition at the grade level. The robustness of our baseline findings in the subsample of schools with a single classroom per grade further alleviates concerns. In addition, our balancing and falsification tests suggest that our estimates are not contaminated by selection.

We also find that native students with low parental education are mostly impacted by the concentration of recent immigrant children in their classroom. One potential explanation for this finding is that natives from disadvantaged families are lacking resources at home to substitute for classroom instruction, which is affected by the presence of recent immigrants. This is however more concerning, as native students from low socio-economic backgrounds are typically the ones exposed to a higher concentration of immigrant peers, which tend to concentrate in more disadvantaged neighborhoods and schools.

Overall, our results suggest that policies putting in place integration programs for recently arrived migrant students could be useful to mitigate those effects, particularly in schools where native and immigrant children disproportionally come from disadvantaged families. Because of the similarities shared by the migration context in the Netherlands with other countries, particularly the predominance of migrants with low socio-economic backgrounds, we believe our findings are of relevance beyond the Dutch context.

References

J. D. Angrist and K. Lang. Does school integration generate peer effects? Evidence from Boston's Metco program. *American Economic Review*, 94(5):1613–1634, 2004. doi:10.1257/0002828043052169.

J. D. Angrist and V. Lavy. Using Maimonides' rule to estimate the effect of class size on scholastic achievement. *Quarterly Journal of Economics*, 114(2):533–575, 1999. doi:10.1162/003355399556061.

R. M. Ballatore, M. Fort, and A. Ichino. The tower of Babel in the classroom: Immigrants and natives in Italian schools. CEPR Discussion Papers 10341, 2015.

A. Bohlmark. Age at immigration and school performance: A siblings analysis using swedish register data. *Labour Economics*, 15(6):739–757, 2008. doi:10.1016/j.labeco.2007.12.004.

T. Brodaty. Peer effects in education: A literature review. *Revue d'économie politique*, 120(5): 739–757, 2010. doi:10.3917/redp.205.0739.

G. Brunello and L. Rocco. The effect of immigration on the school performance of natives: Cross country evidence using PISA test scores. *Economics of Education Review*, 32(1):234–246, 2013. doi:10.1016/j.econedurev.2012.10.006.

S. Carrell, R. Fullerton, and J. West. Does your cohort matter? Measuring peer effects in college achievement. *Journal of Labor Economics*, 27(3):439–464, 2009. doi:10.1086/600143.

K.E. Cortes. The effects of age at arrival and enclave schools on the academic performance of immigrant children. *Economics of Education Review*, 25:121–132, 2006. doi:10.1257/aer.97.2.31.

F. Cunha and J. Heckman. The technology of skill formation. *American Economic Review*, 97(2): 31–47, 2007. doi:10.1257/aer.97.2.31.

E. Duflo, P. Dupas, and M. Kremer. Peer effects, teacher incentives, and the impact of tracking: Evidence from a randomized evaluation in Kenya. *American Economic Review*, 101(5):1739–1774, 2011. doi:10.1257/aer.101.5.1739.

D. Epple and R. Romano. Peer effects in education: Survey of the theory and evidence. In *Handbook of Social Economics*, volume 1, pages 1053–1063. Elsevier BV, 2011. doi:10.1016/B978-0-444-53707-2.00003-7.

C. Geay, S. McNally, and S. Telhaj. Non-native speakers of English in the classroom: What are the effects on pupil performance? *Economic Journal*, 123(570):F281–F307, 2013. doi:10.1111/ecoj.12054.

E. D. Gould, L. Lavy, and M.D. Paserman. Does immigration affect the long-term educational outcomes of natives? Quasi-experimental evidence. *Economic Journal*, 119(540):1243–1269, 2009. doi:10.1111/j.1468-0297.2009.02271.x.

E. A. Hanushek, J. F. Kain, J. M. Markman, and S. G. Rivkin. Does peer ability affect student achievement? *Journal of Applied Econometrics*, 18(5):527–544, 2003. doi:10.1002/jae.741.

E. A. Hanushek, J. F. Kain, and S. G. Rivkin. New evidence about Brown v Board of Education: The complex effects of school racial composition on achievement. *Journal of Labor Economics*, 27(3):349–383, 2004. doi:10.1086/600386.

C. Hoxby. Peer effects in the classroom: Learning from gender and race variation. NBER Working Paper 7867, 2000.

P. Jensen and A. Rasmussen. The effect of immigrant concentration in schools on native and immigrant children's reading and math skills. *Economics of Education Review*, 30(6):1503–1515, 2011. doi:10.1016/j.econedurev.2011.08.002.

V. Lavy and A. Schlosser. Mechanisms and impacts of gender peer effects at school. *American Economic Journal: Applied Economics*, 3(2):1–33, 2011. doi:10.1257/app.3.2.1.

E. P. Lazear. Educational production. *Quarterly Journal of Economics*, 116(3):777–803, 2001. doi:10.1162/00335530152466232.

C. Manski. Identification of endogenous social effects: The reflection problem. *The Review of Economic Studies*, 60(3):531–542, 1993. doi:10.2307/2298123.

A. Ohinata and J. C. van Ours. Young immigrant children and their educational attainment. *Economics Letters*, 116(3):288–290, 2012. doi:10.1016/j.econlet.2012.03.020.

A. Ohinata and J. C. van Ours. How immigrant children affect the academic achievement of native Dutch children. *The Economic Journal*, 123(570):F308–F331, 2013. doi:10.1111/ecoj.12052.

A. Ohinata and J. C. van Ours. Quantile peer effects of immigrant children at primary schools. *Labour*, 30:135–157, 2016. doi:10.1111/ecoj.12052.

B. Sacerdote. Peer effects with random assignment: Results for Dartmouth roommates. *Quarterly Journal of Economics*, 116(2):681–704, 2001. doi:10.1162/00335530151144131.

N. Schneeweis. Immigrant concentration in schools: Consequences for native and migrant students. *Labor Economics*, 35(C):63–76, 2015. doi:10.1016/j.labeco.2015.03.004.

A. Sojournier. Identification of peer effects with missing peer data: Evidence from project STAR. *Economic Journal*, 123(569):574–605, 2013. doi:10.1111/j.1468-0297.2012.02559.x.

M. Tornello. Peer effects of non-native students on natives' educational outcomes: Mechanisms and evidence. *Empirical Economics*, 51(1):383–414, 2016. doi:10.1007/s00181-015-0995-y.

J. C. van Ours and J. Veenman. Age at immigration and educational attainment of young migrants. *Economics Letters*, 90(3):288–290, 2006. doi:10.1016/j.econlet.2005.08.013.

Tables

Table 1: Background characteristics and outcomes of immigrant and native students

		Immigrants			
	Native Dutch	All	Turkish/ Moroccan	Former colonies	Other immigrants
% of students by parental education					
Primary	15.23	43.88	67.41	25.06	32.23
Lower secondary	38.41	25.79	18.17	47.29	26.61
Upper secondary	28.37	16.66	10.96	20.34	20.96
University	17.99	13.67	3.45	7.29	20.09
Total	100	100	100	100	100
Average test score – Dutch language					
Father's education: primary	43.53	41.21	40.18	42.55	42.04
	(10.11)	(10.37)	(10.13)	(10.23)	(10.83)
Father's education: lower secondary	49.51	44.52	41.09	44.18	44.35
	(9.43)	(10.57)	(10.11)	(10.24)	(10.50)
Father's education: upper secondary	52.66	46.89	42.96	45.36	45.97
	(8.67)	(10.54)	(10.46)	(10.33)	(10.48)
Father's education: university	55.10	48.84	46.07	48.99	47.20
	(7.99)	(10.53)	(9.34)	(10.09)	(10.37)
All students	50.46	44.94	40.85	44.23	43.05
	(9.79)	(10.89)	(10.19)	(10.36)	(10.58)
Average test score – mathematics					
Father's education: primary	45.74	44.70	44.34	43.43	45.26
	(10.36)	(10.84)	(10.53)	(10.62)	(10.79)
Father's education: lower secondary	49.12	46.26	45.21	44.11	46.65
	(9.86)	(10.53)	(10.21)	(10.71)	(11.13)
Father's education: upper secondary	52.05	48.35	47.41	45.34	47.96
	(9.01)	(10.16)	(9.77)	(10.23)	(10.58)
Father's education: university	54.34	50.50	50.05	47.69	49.71
	(8.31)	(10.22)	(10.05)	(10.11)	(10.16)
All students	50.29	46.69	45.01	44.21	46.89
	(9.88)	(10.76)	(10.44)	(10.38)	(10.23)
Number of students	347,875	22,450	5,917	1,678	14,855

Note. Individual raw test scores were standardized to have a mean of 50 and a standard deviation of 10. The upper panel reports the distribution of students by parental education, for each subgroup. Figures in the top panel read: 3.45% of Turkish/Moroccan immigrant students have a father that completed higher education. The middle and bottom panels show the average test scores for each subgroup, by level of parental education. Figures in the middle and bottom panels read: Dutch students whose father has primary education have an average verbal test score of 43.53.

Table 2: Summary statistics - aggregate statistics at the grade level

		Percentage of immigrants in the grade				
	All	No immigrant	0-10	10-20	20-50	50+
Grade-level characteristics						
Number of students in the grade	26.34	22.03	29.79	23.11	21.96	29.24
	(13.02)	(12.36)	(14.11)	(11.23)	(11.49)	(15.82)
Fraction of immigrant students	0.063	-	0.053	0.137	0.279	0.820
	(0.133)	-	(0.021)	(0.027)	(0.072)	(0.180)
Share of natives with low parental education	0.162	0.111	0.164	0.261	0.343	0.378
	(0.23)	(0.19)	(0.21)	(0.26)	(0.28)	(0.36)
Average test score in Dutch language						
All students	49.89	51.13	49.77	47.37	45.72	45.93
	(5.37)	(5.01)	(4.91)	(5.46)	(5.41)	(5.65)
Immigrant students	44.93	-	45.73	44.10	43.24	45.57
	(9.13)	-	(10.12)	(8.21)	(6.86)	(6.09)
Native students	50.14	51.22	50.81	47.89	46.63	44.87
	(5.42)	(5.04)	(4.97)	(5.71)	(5.79)	(8.15)
Average test score in mathematics						
All students	49.88	50.76	49.76	48.08	46.99	47.41
	(4.94)	(4.82)	(4.58)	(4.95)	(5.12)	(4.76)
Immigrant students	46.86	-	47.30	46.37	45.94	47.12
	(9.03)	-	(10.05)	(8.10)	(7.10)	(5.07)
Natives	50.00	50.81	49.89	48.35	47.34	46.99
	(5.04)	(4.87)	(4.68)	(5.16)	(5.33)	(6.75)
Number of grade-level observations	12,053	6,522	3,403	1,322	686	120

Note. Reported statistics were aggregated at the grade level within a school. Standard deviations at the grade level are reported in parentheses. Natives with low parental education are defined as having a father that did not complete upper secondary education.

Table 3: Balancing tests for the validity of the identification strategy

Dependent variable: % of immigrants in the grade	Ordinary Least Squares (1)	School fixed effects (2)	School-by-year fixed effects (3)	School-by-year fixed effects + linear trend (4)
% of natives whose father has primary education	0.147***	-0.010	-0.001	-0.004
	(0.013)	(0.007)	(0.006)	(0.005)
% of natives whose father has lower secondary education	-0.012	0.002	0.001	0.002
	(0.0122)	(0.005)	(0.004)	(0.004)
% of natives whose father has upper secondary education	-0.132***	-0.012*	-0.001	-0.001
	(0.013)	(0.006)	(0.006)	(0.004)
% of natives whose father has university education	-0.092***	0.014	-0.004	0.001
	(0.012)	(0.008)	(0.006)	(0.005)
Fraction of female students	0.04**	0.014	-0.002	0.003
	(0.014)	(0.009)	(0.005)	(0.005)
Fraction of natives from disadvantaged families	0.105***	0.010	0.006	0.002
	(0.010)	(0.008)	(0.006)	(0.004)
Average class size	-0.092***	-0.007	-0.010	-0.002
	(0.010)	(0.010)	(0.010)	(0.008)
Total number of students in the grade	0.038***	-0.018**	-0.013	-0.002
	(0.009)	(0.008)	(0.017)	(0.007)
Fraction of students that repeated a grade	0.102***	-0.012	-0.007	0.001
	(0.014)	(0.009)	(0.006)	(0.004)
Teacher's years of experience	0.067***	-0.013	-0.006	0.001
	(0.011)	(0.010)	(0.008)	(0.007)
Number of grade-level observations	12,053	12,053	12,053	12,053

Notes. ***: significant at the 1% level, **: significant at the 5% level, *: significant at the 10% level. Each row reports estimates from separate regressions of the percentage of immigrant students in the school grade on the corresponding explanatory variable. Robust standard errors clustered at the school level are reported in parentheses. All regressions include grade dummies.

Table 4: Effects of the share of immigrant classmates in the grade

	Natives' language score		Natives' math score	
	(1)	(2)	(3)	(4)
Treatment 1: Share of immigrants	-0.776	-0.751	-0.185	-0.191
	(0.697)	(0.769)	(0.701)	(0.706)
Treatment 2: Share of recent immigrants	-3.08***	-2.88***	-1.55	-1.49
	(0.981)	(0.977)	(0.913)	(0.906)
Treatment 3: Share of other immigrants	0.145	0.168	0.012	-0.027
	(0.688)	(0.676)	(0.691)	(0.653)
Enrollment (2nd polyn.)		✓		✓
Grade-level controls		✓		✓
Grade effects	✓	✓	✓	✓
School-by-year effects	✓	✓	✓	✓
Number of grade-level observations	12,053	12,053	12,053	12,053

Notes. ***: significant at the 1% level, **: significant at the 5% level, *: significant at the 10% level. Robust standard errors clustered at the school level are reported in parentheses. Each row reports the coefficients from separate regressions of the effect of the corresponding treatment on natives' average test scores. Controls for grade-level characteristics include: the share of students by level of parental education, the share of female students in the grade, the share of disadvantaged students according to the Dutch weighting system, the share of students that repeated a grade, the average class size in the grade, teacher's years of experience.

Table 5: Heterogeneous treatment effects by natives' parental education

	Natives with high parental education				Natives with low parental education			
	Language score		Math score		Language score		Math score	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A:								
Share of recent	-1.131	-1.184	-0.935	-0.620	-3.34**	-3.35**	-1.13	-1.12
immigrants in grade	(1.291)	(1.223)	(1.178)	(1.171)	(1.457)	(1.446)	(1.402)	(1.379)
Enrollment (2nd polyn.)		✓		✓		✓		✓
Grade-level controls		✓		✓		✓		✓
Grade effects	✓	✓	✓	✓	✓	✓	✓	✓
School-by-year effects	✓	✓	✓	✓	✓	✓	✓	✓
N. of grade-level observations	11,062	11,062	11,062	11,062	11,664	11,664	11,664	11,664
Panel B:								
Share of other	0.202	0.156	0.090	0.118	-1.503	-1.462	-0.608	0.587
immigrants in grade	(1.425)	(1.375)	(1.502)	(1.563)	(1.657)	(1.276)	(1.643)	(1.576)
Enrollment (2nd polyn.)		✓		✓		✓		✓
Grade-level controls		✓		✓		✓		✓
Grade effects	✓	✓	✓	✓	✓	✓	✓	✓
School-by-year effects	✓	✓	✓	✓	✓	✓	✓	✓
N. of grade-level observations	11,062	11,062	11,062	11,062	11,664	11,664	11,664	11,664

Notes. ***: significant at the 1% level, **: significant at the 5% level, *: significant at the 10% level. Robust standard errors clustered at the school level are reported in parentheses. Controls for mean characteristics at the grade level include: the share of students by level of parental education, the share of female students in the grade, the share of disadvantaged students according to the Dutch weighting system, the share of students that repeated a grade, the average class size in the grade, teacher's years of experience. Low parental education refers to having a father that did not complete upper secondary education while high levels of parental education are defined as having a father that completed upper secondary education or more.

Table 6: Falsification tests – placebo regressions

	Natives' language score		Natives' math score	
	(1)	(2)	(3)	(4)
Treatment variable:				
Actual treatment: Share of recent immigrants in grade	-3.08***	-2.88***	-1.55	-1.49
	(0.981)	(0.977)	(0.913)	(0.906)
Placebo 1: Share of recent immigrants in next grade	-0.361	0.120	0.038	0.179
	(1.241)	(0.112)	(1.201)	(1.192)
Placebo 2: Share of recent immigrants in previous grade	0.526	0.534	-0.472	-0.294
	(1.189)	(1.181)	(1.092)	(1.071)
Enrollment (2nd polyn.)		✓		✓
Grade-level controls		✓		✓
Grade effects	✓	✓	✓	✓
School-by-year effects	✓	✓	✓	✓
Number of grade-level observations	12,053	12,053	12,053	12,053

Notes. ***: significant at the 1% level, **: significant at the 5% level, *: significant at the 10% level. Robust standard errors clustered at the school level are reported in parentheses. Each row reports the coefficients from separate regressions of the effect of the corresponding treatment on natives' average test scores. Controls for mean characteristics at the grade level include: the share of students by level of parental education, the share of female students in the grade, the share of disadvantaged students according to the Dutch weighting system, the share of students that repeated a grade, the average class size in the grade, teacher's years of experience

Table 7: Treatment effects using variation in immigrant concentration in grade 4 within the same school across years

	Natives' language score		Natives' math score	
	(1)	(2)	(3)	(4)
Treatment 1: Share of immigrants	-1.102	-1.200	-0.985	-0.957
	(1.336)	(1.273)	(1.400)	(1.376)
Treatment 2: Share of recent immigrants	-3.78**	-3.72**	-1.75	-1.67
	(1.301)	(1.210)	(1.415)	(1.302)
Treatment 3: Share of other immigrants	0.145	0.168	0.142	0.128
	(1.347)	(1.267)	(1.368)	(1.333)
Enrollment (2nd polyn.)		✓		✓
Year effects	✓	✓	✓	✓
School-by-grade effects	✓	✓	✓	✓
Number of grade-level observations	3,015	3,015	3,015	3,015

Notes. ***: significant at the 1% level, **: significant at the 5% level, *: significant at the 10% level. Robust standard errors clustered at the school level are reported in parentheses. Each row reports the coefficients from separate regressions of the effect of the corresponding treatment on natives' average test scores. Controls for mean characteristics at the grade level include: the share of students by level of parental education, the share of female students in the grade, the share of disadvantaged students according to the Dutch weighting system, the share of students that repeated a grade, the average class size in the grade, teacher's years of experience.

Table 8: Linear treatment effect in full sample and restricted sample

	Full sample				Schools with a single class per grade			
	Natives' language score		Natives' math score		Natives' language score		Natives' math score	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Share of recent migrants in grade	-3.08***	-2.88***	-1.55	-1.49	-3.60***	-3.34***	-1.48	-0.980
	(0.981)	(0.977)	(0.913)	(0.906)	(1.28)	(1.38)	(1.27)	(1.19)
Enrollment (2nd polyn.)		✓		✓		✓		✓
Grade-level controls		✓		✓		✓		✓
Grade effects	✓	✓	✓	✓	✓	✓	✓	✓
School-by-year effects	✓	✓	✓	✓	✓	✓	✓	✓
Number of observations	12,053	12,053	12,053	12,053	8,188	8,188	8,188	8,188

Notes. ***: significant at the 1% level, **: significant at the 5% level, *: significant at the 10% level. Robust standard errors clustered at the school level are reported in parentheses. Controls for grade-level characteristics include: the share of students by level of parental education, the share of female students in the grade, the share of disadvantaged students according to the Dutch weighting system, the share of students that repeated a grade, the average class size in the grade, teacher's years of experience.

Appendix

Table A1: Balancing tests for the share of recent immigrants in the grade

Dependent variable: % of recent immigrants in grade	Ordinary Least Squares	School fixed effects	School-by-year fixed effects	School-by-year fixed effects + linear trend
% of natives whose father has primary education	0.162***	-0.009	-0.002	-0.002
	(0.014)	(0.007)	(0.008)	(0.007)
% of natives whose father has lower secondary education	-0.011	0.003	-0.001	0.001
	(0.0122)	(0.005)	(0.006)	(0.006)
% of natives whose father has upper secondary education	-0.145***	-0.015*	0.001	-0.001
	(0.013)	(0.007)	(0.007)	(0.005)
% of natives whose father has university education	-0.113***	0.017*	-0.005	-0.001
	(0.012)	(0.008)	(0.008)	(0.006)
Fraction of female students	0.051**	0.014	-0.002	-0.003
	(0.014)	(0.009)	(0.006)	(0.005)
Fraction of natives from disadvantaged families	0.102***	0.012	0.003	0.001
	(0.013)	(0.008)	(0.005)	(0.004)
Average class size	-0.071***	-0.012	-0.010	-0.002
	(0.012)	(0.010)	(0.011)	(0.008)
Total number of students in the grade	0.041***	-0.020**	-0.011	-0.001
	(0.011)	(0.008)	(0.016)	(0.007)
Fraction of students that repeated a grade	0.131***	-0.014	-0.005	-0.002
	(0.014)	(0.009)	(0.008)	(0.007)
Teacher's years of experience	0.098***	0.011	0.003	0.001
	(0.019)	(0.012)	(0.007)	(0.006)
Number of grade-level observations	12,053	12,053	12,053	12,053

Notes. ***: significant at the 1% level, **: significant at the 5% level, *: significant at the 10% level. Each row reports estimates from separate regressions of the percentage of recent immigrant students in the grade on the corresponding explanatory variable. Robust standard errors clustered at the school level are reported in parentheses. All regressions include grade dummies.

Table A2: Treatment effects using variation in immigrant concentration in grade 6 and 8 within the same school across years

	Natives' language score		Natives' math score	
	(1)	(2)	(3)	(4)
Panel A: Grade 6				
Treatment 1: Share of immigrants	-0.912	-0.980	-0.782	-0.701
	(1.401)	(1.365)	(1.422)	(1.379)
Treatment 2: Share of recent immigrants	-4.35**	-4.26**	-1.98	-1.77
	(1.499)	(1.475)	(1.482)	(1.379)
Treatment 3: Share of other immigrants	0.145	0.168	0.142	0.128
	(1.347)	(1.267)	(1.368)	(1.333)
Enrollment (2nd polyn.)		✓		✓
Year effects	✓	✓	✓	✓
School-fixed effects	✓	✓	✓	✓
Number of grade-level observations	3,017	3,017	3,017	3,017
Panel A: Grade 8				
Treatment 1: Share of immigrants	0.605	0.555	-0.654	-0.602
	(1.428)	(1.400)	(1.472)	(1.424)
Treatment 2: Share of recent immigrants	-3.33*	-3.48**	-2.37	-2.24
	(1.402)	(1.321)	(1.522)	(1.504)
Treatment 3: Share of other immigrants	0.801	0.777	0.405	0.422
	(1.455)	(1.423)	(1.428)	(1.400)
Enrollment (2nd polyn.)		✓		✓
Year effects	✓	✓	✓	✓
School-fixed effects	✓	✓	✓	✓
Number of grade-level observations	3,014	3,014	3,014	3,014

Notes. ***: significant at the 1% level, **: significant at the 5% level, *: significant at the 10% level. Robust standard errors clustered at the school level are reported in parentheses. Each row reports the coefficients from separate regressions of the effect of the corresponding treatment on natives' average test scores. Controls for mean characteristics at the grade level include: the share of students by level of parental education, the share of female students in the grade, the share of disadvantaged students according to the Dutch weighting system, the share of students that repeated a grade, the average class size in the grade, teacher's years of experience.