

The Effect of Immigration on Natives' School Achievement

Does Length of Stay in the Host Country Matter?

Laurent Bossavie



WORLD BANK GROUP

Social Protection and Jobs Global Practice

June 2018

Abstract

Using a rich data set of primary school students, this paper estimates the effects of immigrant concentration in the classroom on the academic achievement of natives. In contrast with previous contributions, it exploits rare information on age-at-migration to estimate separate spillover effects by duration of stay of immigrant classmates. To identify treatment effects, it uses cohort-by-cohort deviations in immigrant concentration within schools combined with attractive features of the Dutch school system. Overall, the paper finds no effect of the concentration of immigrant students on natives' test

scores. However, although immigrant students who have been in the country for some time have virtually no effect on natives, the analysis finds a small negative effect of recent immigrants in the classroom on natives' test scores. The effect is significant only for language test scores, but insignificant for mathematics test scores. When significant, effect sizes are quite small compared to other educational interventions and classroom peer effects estimated in other contexts..

This paper is a product of the Social Protection and Jobs Global Practice. It is part of a larger effort by the World Bank to provide open access to its research and make a contribution to development policy discussions around the world. Policy Research Working Papers are also posted on the Web at <http://www.worldbank.org/research>. The author may be contacted at lbossavie@worldbank.org.

The Policy Research Working Paper Series disseminates the findings of work in progress to encourage the exchange of ideas about development issues. An objective of the series is to get the findings out quickly, even if the presentations are less than fully polished. The papers carry the names of the authors and should be cited accordingly. The findings, interpretations, and conclusions expressed in this paper are entirely those of the authors. They do not necessarily represent the views of the International Bank for Reconstruction and Development/World Bank and its affiliated organizations, or those of the Executive Directors of the World Bank or the governments they represent.

The Effect of Immigration on Natives' School Achievement:
Does Length of Stay in the Host Country Matter?*

Laurent Bossavie[†]

Keywords: Immigration, education, peer effects

JEL classification: I21, J15

* I am extremely grateful to the Data Archiving and Network Services (DANS) in the Netherlands, in particular to Hernan Vierke, for granting me the access to the various waves of the PRIMA dataset. I am also indebted towards Andrea Ichino and Luigo Guiso for constant advice and discussions at the earlier stages of the paper. All remaining errors are mine.

[†] Current address: The World Bank, 1818 H Street NW, 20433 Washington DC, USA. E-mail: lbossavie@worldbank.org. Tel: 202-751-6478. An earlier version of this paper was written while the author was a PhD student at the Department of Economics of the European University Institute (EUI). The findings, interpretations and conclusions expressed in this paper are entirely those of the author. They do not necessarily reflect the views of the European University Institute or those of the World Bank Group.

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 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 30 percent the share of foreign-born students in public school classrooms. Such measures, however, are largely motivated by anecdotal evidence of disruption rather than clear-cut results of rigorous econometric 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

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

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 studies is that they typically treat immigrant children as an homogeneous group. In particular, they do not take into account the duration of stay of foreign-born children in the host country when estimating peer effects. There are however 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 of immigrant students in the host country has been positively linked to their test score performance.² Immigrant students who 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 the move to a new country. They may therefore require greater attention from teachers compared to immigrant children who arrived to the country at an earlier age, which could affect instruction to the entire classroom.

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 duration of stay of immigrant classmates in the Netherlands, it separately estimates the impact of foreign-born peers who recently arrived to the Netherlands from those who arrived in

²Contributions such as Ohinata and van Ours (2012) for the Netherlands have reported a statistically significant and positive association between duration of stay in the host country and scholastic achievement of foreign-born students.

the country at an earlier age. Previous work on the topic typically does not make this distinction when estimating the effects of immigrant concentration in the classroom.³

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.⁴ On the other hand, using grade-level peer composition can estimate peer effects imprecisely or even lead to a downward bias, if most learning spillovers 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. We present a battery of tests and robustness checks 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

³Although this is not the main focus of their paper, the only exception is Branden et al. (2016) in the Swedish context.

⁴One recent exception is Ballatore et al. (2015) who attempt to account for the endogeneity of classroom formation to identify the effect of immigrant classmates.

host country at those ages than older students. One could therefore expect greater disparities with 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 arithmetic in fostering further skills and shaping labor market outcomes.⁵

Our results suggest that the impact of immigrant concentration on natives' test scores is heterogeneous, both in the type of immigrants who are part of the treatment, but also in the type of natives who 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 who have been in the country for a short period. The effect size is however quite small in magnitude, and statistically significant only for scholastic achievement in Dutch language. Furthermore, we report that native students with high parental education are not affected by the concentration of immigrant classmates in their classroom, even if those are recent migrants.

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 and discusses our main results, while Section 7 performs placebo tests and sensitivity checks. Section 8 concludes.

⁵See Heckman and Cunha (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 attempted to estimate the effects of classroom composition on individual school performance.⁶

Evidence on the impact of immigrant classmates on natives' scholastic achievement is more scarce. Diette and Oyelere (2017) and Neymotin (2009) are two studies that estimate the impact of immigration on natives' school outcomes in the US context. Diette and Oyelere (2017) look at the effects of Limited English (LE) classmates on natives' test scores in North Carolina.⁷ Using a school-by-year fixed effect estimator, they find no evidence of negative peer effects of Limited English ability students on females and white students, but note small negative effects on average on males and black students. They also report that an increase in the share of Latin American students per se does not create negative peer effects on native students' achievement, but that the limited English language skills of some of these students appears to be generating small, negative peer effects on natives. Neymotin (2009) investigates the effect of immigration on different natives' outcomes in California and Texas, namely SAT scores and college application patterns. Using a

⁶Epple and Romano (2011) or Brodaty (2010) provide a literature review of applied work estimating peer effects in the classroom.

⁷Limited English students refers to students with limited English proficiency.

set of empirical strategies to account for selection, she finds that 1990s immigration did not harm, and possibly benefited the student outcomes of U.S. citizens.

Despite the importance of this question 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 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), Tonello (2016) and Branden et al. (2016).⁸ While Ohinata and van Ours (2013, 2016), Geay et al. (2013) and Schneeweis (2015) report no effect on natives' test scores, other studies find statistically significant negative impacts.

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

⁸Outside 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 (TIMSS) 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 for immigrant student 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. Finally, Branden et al. (2016) use Swedish population registry data to investigate this question. Using a two-way family and school fixed effect estimator, they find no effect of

the share of immigrant students in the school on natives' grades, but report a small negative effect on levels of eligibility for upper secondary school. Although it is not the focus of the paper, it is to the best of our knowledge the only study prior to this paper that estimates distinct effects of immigrant concentration by age at migration.

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 who 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 was populated by a population of 1.77 million immigrants, representing around 11 percent of the country's population.⁹ As in most European countries, the majority of immigrants residing in the Netherlands come from lower-income countries. First, large groups of immigrants came from former Dutch colonies, mainly Indonesia, Suriname and the Dutch Antilles, starting from the 1940s. The independence of Indonesia in 1949 and of Suriname in 1975 led to large influxes of migrants from these two countries. Another large inflow occurred around 1980

⁹This figure includes both first and second generation migrants.

when the mandatory entry visa for Surinamese was introduced, since many feared that entry to the Netherlands would become more restricted (Ohinata and van Ours (2013)). Immigrants from ex-Dutch colonies had mostly a good command of the Dutch language when they entered the country, and are comparatively well-educated within school systems modeled on the Netherlands.

Another immigration wave consisting of low-skilled guest workers, primarily from Turkey and Morocco, entered the Netherlands in the 1960s. This second immigration wave was largely driven by increased demand for low-skilled labor. As a result, the large majority of Turkish and Moroccan immigrants populating the Netherlands are from families of lower socio-economic background compared to native Dutch or migrants from the ex-Dutch colonies.

Although these large migration waves took place several decades ago, the geographical composition of first-generation migrants in the Netherlands over the period studied by this paper strongly reflects those earlier migration waves. After the recruitment of guest workers stopped in the early 1970s, immigration in particular from Morocco and Turkey continued due to family formation and unification (van Ours and Veenman (2006)). There has also been a continuous flow of immigrants from the Netherlands Antilles over the past decades. In addition to these traditional groups, the Netherlands started receiving smaller immigrant groups from Iraq, Afghanistan or Iran, mostly asylum seekers, starting from the 1990s. 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.

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 1.7.¹⁰ Finally, children of immigrant parents with low education receive the highest weight of 1.9.

Practically, the formula used to calculate school funding based on student weights allocates funding proportionally to the average weight of students in the school.¹¹ Funding, however, does not increase up to an average student weight of 1.09. Once the average school weight has passed the 1.09 threshold, an increase in the average weight of students by 0.1 increases funding per student by 10%. As an illustration, an increase by one standard deviation (13 percentage points) in the share of immigrant students in a school with average native composition (a mean native student weight of 1.28) would increase school funding per student by about 6%, according to the formula. Schools have some amount of freedom in deploying these extra resources, although they must primarily be used for personnel. As a result, schools typically use those resources to reduce class size, hire more experienced teachers, offer remedial teaching or appoint classroom assistants (Ladd and Fiske (2011)). The additional funding can also be used to introduce more specific measures, such as school-wide language policies or reception facilities for newcomers.

¹⁰The term Bargee is used to describe children who live on ships with their parents on the waterways of Europe, and specifically used in the Netherlands. Children of itinerant parents refer to children of parents who live in mobile homes or caravans.

¹¹For further detail about the exact formula used to determine funding per pupil, see Ladd and Fiske (2011) or Dobbelsteen et al. (2002).

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.¹² 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.¹³

The data collected in PRIMA are 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 who were born abroad, and does not

¹²The 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.

¹³We 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.

include second generation migrants. Both Western and non-Western immigrants are included in our definition of migrants, although the large majority immigrant students in the PRIMA sample are of non-Western origin (95%).

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.

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 low levels of parental education compared to native students, as it is the case in most European countries. In line with historical patterns of migration to the Netherlands, the largest groups of first-generation immigrants among our sample of primary school students are Turkish/Moroccans (27%) followed by students from the ex-Dutch colonies (7.5%). 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. Among the Moroccan and Turkish students, 67 percent have a father who 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 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 different 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 who did not complete upper secondary education ranges from 40 percent in grades with no immigrant to close to 60 percent in grades with a proportion of immigrant students higher than 20 percent. 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.

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 fewer immigrants in their classroom even if the treatment intensity was the same for both types of native students, 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 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 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. As a result, using grade-level measures of immigrant concentration

instead of classroom-level measures may lead to more imprecise peer effect estimates (Gould et al. (2009)).

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.

Given this feature of the Dutch context, the correlation between the grade-level and classroom-level share of immigrants is very high in our sample (0.92). The standard trade-off between grade and classroom-level measures of peers is therefore not restrictive in our context.¹⁴ The main conceptual argument for using grade-level instead of classroom-level observation is the possible manipulation of classroom composition by teachers and principals. In that regard, the decentralized nature of the Dutch primary school system leaves room to principals and parents to manipulate classroom formation, as pointed out by Ammermueller and Pischke (2009) or Ohinata and van Ours (2013). Although schools with multiple classrooms per grade constitute only 30% of our

¹⁴See Gould et al. (2009), Lavy et al. (2012b), Hanushek et al. (2009), or Carrell et al. (2009), among others, for discussions on using grade-level versus classroom-level peer composition

sample, we tested for the non-random allocation of students to classes, using the Pearson-X2 test.¹⁵ The null hypothesis that immigrant students are randomly allocated to classrooms in our subsample of schools with multiple classrooms per grade was marginally rejected. For this reason, the use of grade-level measures of peers appears to offer a slightly cleaner source of variation for the treatment of interest in our context.

A great deal of selection into peer groups occurs between schools. The inclusion of school fixed effects accounts for the most obvious source of student sorting between schools. This selection is likely to be strong in the Netherlands, where a free school choice policy applies. However, there might also be school-specific time varying factors that affect 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, specificities of the Dutch primary school system also require controlling for school-by-year fixed effects. As outlined in section 3.2, the primary school budget in the Dutch context is directly tied to the school socio-economic composition, which can vary across years. As school resources have been shown to affect students' outcomes and are directly affected by the share of immigrant students in a school in a given year, controlling for year-specific school effects is essential. We therefore add a full set of school-year fixed effects γ_{sy} to our specification.

Since test scores of students in the same grade are likely to be correlated which would 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

¹⁵The Pearson X2-square test, also used by Ammermueller and Pischke (2009), asks whether there are more students with a given characteristic immigrant background in my case - in a particular class than is consistent with independence, given the number of students in the school.

test score of native students. Using our panel of schools observed in four different grades over several years, we estimate β , the effect of immigrant concentration in the grade on natives' test scores, from the following reduced-form baseline equation:

$$\bar{Y}_{sgy} = \alpha_g + \gamma_{sy} + \beta I_{sgy} + \rho \bar{X}_{sgy} + \varepsilon_{sgy}, \quad (1)$$

where s denotes the school, y denotes the year, and g the grade. \bar{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. \bar{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 estimating β , which is identified from variations in the proportion of immigrant students across grades within the same school, observed in a given year. The identifying assumption is that changes in the share of immigrant students across grades 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 across grades, there exists cohort-by-cohort variations that are purely driven by exogenous factors.

Even after controlling for school-by-year fixed effects, one might still be concerned that variation in immigrant concentration across grades within schools may be correlated with unobservable cohort factors. Students in different grades within the same school started primary school in different years. The youngest cohort we observe in a given year (grade 2) entered the school six years later than the oldest cohort we observe (grade 8). Although six years is a relatively short

time-span to observe trends in neighborhood and school composition, secular trends correlated with immigrant concentration cannot be completely ruled out. To alleviate this concern, we estimate a second equation, where a full set of school-specific linear trends σ_{sy} are added to Equation 1. This approach is similar in spirit to that of Lavy and Schlosser (2011), Lavy et al. (2012a) or Schneeweis (2015), adapted to our identification strategy which uses school-by-grade fixed effects. The reduced-form equation to estimate the effect of immigrant concentration in the grade becomes:

$$\bar{Y}_{sgy} = \alpha_g + \gamma_{sy} + \sigma_{sy}grade + \beta I_{sgy} + \rho \bar{X}_{sgy} + \varepsilon_{sgy} \quad (2)$$

where the index variable *grade* in Equation (2) takes the value 2, 4, 6 and 8 corresponding to the grade level, and is interacted with school by year dummies to capture school-specific linear trends. In Equation 2, β is identified from the deviations in the proportion of immigrant students in the grade from its linear trend across grades within the same school.

5.3 Evidence for the validity of the identification strategy

A first potential threat to the identification strategy is the fact that families might react to changes in immigration concentration within schools by moving 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 confounding factor is grade retention which is relatively common in the Netherlands, and may therefore potentially lead to non-random variation of native students' characteristics across grades that have different concentrations of immigrants students. In our sample of Dutch primary schools, 14.4% of students in grade 2 to 8 have repeated at least one grade. The share of repeaters is increasing with grade level and is 9% for students in grade 2 and 17.3% in grade 8.¹⁶ In this 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.

Finally, a last potential threat to identification is the non-random allocation of school resources to grades with more migrants. As schools with a higher share of migrants receive more funding per student, one could suspect that principals may assign this extra money disproportionately to grades with a greater share of immigrants. An important institutional feature given our identification strategy, however, is that nothing in the design and implementation of the Dutch program mandates that the extra resources occasioned by the student weights are to be used for the students to whom the weights are attached (Ladd and Fiske (2011)). This implies that the additional resources received by the school do not have to be allocated to grades with a higher share of migrant students, who receive the highest weights. Moreover, the inclusion of a threshold provision in the Dutch additional funding system means that in practice, there are no additional resources for a significant proportion of students who have weights associated with them. However, we investigate this possibility in this

¹⁶We identify grade repeaters as students who are enrolled in a lower grade than predicted by their month and year of birth. In the Netherlands, the school cohort cutoff is set to September 30th. Therefore, a given school cohort consists of all pupils born between October 1 of year t and September 30 of the following year $t+1$. We therefore identify students who are born prior to October 1st of year t as repeaters.

section.

To test for potential non-random variation in immigrant concentration across grades, we regress our treatment variable, i.e. the fraction of immigrant students, on the characteristics of native students in the same grade, as well as measures of grade-level teaching resources such as average class size, teachers' experience and whether the grade has supporting teaching staff. Those grade-level measures of teaching resources allow to test whether schools allocate more resources to grades with more immigrant students, a potential threat to our identification strategy. As detailed in Ladd and Fiske (2011), the large majority of additional funding received based on the number of immigrant children in the grade had to be allocated to personnel. Given these allocation rules, schools with higher levels of funding are expected to have more teaching resources, resulting in lower class size, more experienced teachers (as teachers' salary levels are based on experience), as well as additional support staff.¹⁷ We therefore include these grade-level measures of teaching resources in our balancing tests.

The results of these balancing tests are reported in Table 3. 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 1 percent level. As evidenced earlier, natives with low parental education tend to concentrate in schools with a high fraction of immigrant students. In addition, there is a significant negative association between average class size and the share of immigrants in the grade, and a positive

¹⁷Those patterns have been confirmed empirically by Ladd and Fiske (2011) using administrative data on Dutch primary schools.

association between immigrant concentration in the grade and teachers' years of experience. This confirms that principals use additional resources to reduce class size and hire more experienced teachers at the school level.

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 the total number of students enrolled in the grade, as well as whether the grade has a remedial teacher, which remain statistically significant at the 5% level. Using within-school variation in immigrant concentration therefore substantially alleviates selection issues, although some significant association remains with two of the variables. Once school-fixed effects are accounted for, the association between treatment intensity and grade-level characteristics is brought to virtually zero for the large majority of variables.

Column 3 shows the association between the share of immigrants in the grade and the same set of natives' characteristics and teaching resources 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 correlation with natives' characteristics, which become virtually zero. In addition, there is no remaining association between immigrants in the grade and the total number of students enrolled. There is also no association of our treatment variable with average class size as well as teacher's experience. This finding is inconsistent with principals decreasing class size and allocating more experienced teachers in grades where there are more migrants, within the same school. The correlation with having a teaching aide, having a remedial

teacher, or whether the grade offers subgroup teaching also becomes insignificant.¹⁸

Finally, Column 4 shows the association between grade characteristics and the fraction of immigrants when school linear trends are also controlled for. The magnitude of all correlations are virtually zero and very similar to the school-by-year fixed effect estimates. 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 who 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 and potential non-random allocation of teaching resources. 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.

¹⁸A remedial teacher refers to an additional teacher who works across grades.

6 Results

6.1 Effects of Immigrant Concentration

Row 1 of Table 4 reports the linear effects of the share of immigrants in the grade on the average test score of natives (Treatment 1), obtained by estimating Equation 2.¹⁹ This is the standard treatment effect estimated by the literature looking at the impact of immigrant concentration on natives' test scores. 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 the Netherlands only have one classroom per grade. The estimated effect size is very small: 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, relative 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 if variation in treatment intensity is random. We therefore find no impact of immigrant concentration on natives' achievement when immigrant students are treated as a homogenous group. 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.

¹⁹Estimates that do not control for linear school trends are quantitatively very similar and available upon request.

6.2 Effect of Immigrant Concentration by Duration of Stay

Existing evidence suggests that young immigrant children who have been in the country for a longer period tend to perform better in school compared to immigrant children who have been in the country for a short period (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. We therefore hypothesize that the negative effect of the concentration of immigrant peers who recently migrated might be larger than that of foreign-born students who have been in the host country longer.

To test this hypothesis, we exploit 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 who 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). Rows 2 and 3 of Table 4 report estimates for these two alternative treatment effects. The share of recent immigrants in the grade has a negative and statistically significant effect on natives' verbal test scores, but the estimated effect size remains small in magnitude. According to our estimation, an increase of the share of recent immigrants in the grade by one standard deviation (five percentage points)

reduces natives' average language test score by about 0.03 standard deviation of the average native language test score. 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 Effect of Immigrant Concentration by Geographical Origin

We also investigate the effect of immigrant concentration by country of origin of foreign-born students. There are several reasons to suspect that the effect of immigrant concentration in the classroom may also vary by migrants' country of origin. First of all, country of origin is likely to be associated with different levels of command of the Dutch language upon arrival. Second, migrants from different countries vary in terms of parental and socio-economic background, which could also generate different spillovers on natives.

As mentioned in Section 3.2, migrants from the ex-Dutch colonies would in the majority of cases speak Dutch upon arrival or would at least have been exposed to the language at home. As reported in Table 2, immigrant students from former Dutch colonies perform significantly better on average than Turkish/Moroccan students in Dutch language. The achievement gap remains statistically significant once parental education is accounted for, suggesting that differences in parental education between the two groups is insufficient to explain the gap. This evidence is consistent with the large degree of dissimilarity between Turkish and Dutch or Arabic and Dutch evidenced by the Levenshtein linguistic distance index, from Adsera and Pytlikova (2015).²⁰

²⁰The Levenshtein linguistic distance relies on phonetic dissimilarity of words in two languages, and produces a continuous index that increases with the distance between languages. According to this measure, Turkish and Arabic have a dissimilarity index with respect to Dutch of 102.3 and

Given this pattern and the fact that Turkish/Moroccan immigrants have lower levels of parental education than migrants from other origins (Table 1), we estimate separate spillovers of foreign-born students from Turkey/Morocco versus migrants from other origins. Migrants from origins other than Turkey and Morocco are heterogeneous in their native languages but exhibit notably higher levels of parental education than migrants from those countries, although lower than native Dutch.

The estimates of the treatment effect of immigrants from Moroccan/Turkish origin versus other origins, irrespective of their duration of stay in the Netherlands, are displayed in Table 5. Estimated effects are quite small in magnitude and statistically insignificant, irrespective of the country of origin of migrants. Therefore, while length of stay matters for spillovers between immigrant and native students, country of origin appears to play little role when duration of stay is not taken into account.

We therefore also investigate whether migration duration of stay has different spillover effects on natives, depending on migrant origin. To ensure sufficient variability in the treatment variable, however, we construct an alternative treatment measure where we exclude immigrants from ex-Dutch colonies as well as those that have one Dutch parent from our measure of immigrant concentration.²¹ In other words, we calculate the share of recent immigrants by excluding migrants who are more likely to have been exposed to Dutch prior to migrating, and denote this treatment 100 respectively, compared to a minimum value of 0 (perfect similarity) and a maximum of 106, for a median dissimilarity value of 86 for the 217 countries in the Max Planck Institute sample.

²¹We also attempted to estimate separate spillovers for Turkish/Moroccan migrants who recently migrated as well as recent migrants from other origins. However, distinguishing between both duration of stay and origin of migrants to measure treatment intensity generated large standard errors, due to limited within school variation in the share of foreign-born students who are both recent migrants and from a particular origin. Estimates using these measures of treatment were therefore quite imprecise, and lacked significance for all subgroups mainly due to large standard errors.

”Share of recent immigrants with low prior exposure to Dutch”.

We compare the effect of this alternative treatment to that of the baseline estimate of immigrant concentration by duration of stay in Table 6. As shown in the table, the magnitude of the estimated effect increases from -3.05 to -3.70. This suggests that immigrants who are less likely to be familiar with Dutch upon arrival exert a stronger negative effect on native Dutch students, shortly after their arrival. For migrants who have been staying in the Netherlands for a longer period, however, spillovers are statistically insignificant, even when immigrants had low exposure to Dutch prior to migrating. Those findings are consistent with recent evidence from Diette and Oyelere (2017) or Frattini and Meschi (2017), which suggest that negative spillovers associated with the concentration of immigrant students in the classroom originate from limited language skills of foreign-born students, rather than from classroom ethnic heterogeneity per se.

6.4 Heterogeneous Effects by Natives’ Types

We previously assumed that the effect of immigrant concentration was identical for all natives. However, some of the literature on classroom peer effects suggests that spillovers might be heterogeneous across student types. Hanushek et al. (2003) and Lavy et al. (2012a) find that the test scores of students in the lower end of the ability distribution are more negatively impacted by the presence of low-performing students in their grade. To investigate this possibility in our context, we look at the impact of immigrant concentration on two types of natives by estimating separately the impact on natives with low parental education and high parental education, as proxies for socio-economic status.

We run the same regressions as in Table 4 separately for these two groups of natives. Results

are presented in Table 7. 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 impacted by the presence of recent immigrants.

6.5 Effect sizes

When statistically significant, the negative effects of having a higher concentration of recent immigrants in the classroom are quite small. To give the reader a sense of the magnitude of the effects, it is useful to compare them to well-known educational interventions. One of the most studied educational inputs is class size. The cleanest available evidence so far comes from the Tennessee STAR experiment studied by Krueger (1999) as well as later contributions, summarized in Hanushek (2006). The STAR sample consists of primary school students of similar age as in our sample. Krueger (1999) reports a negative effect size for an increase in class size by one student of about 0.03 standard deviation in test scores. The effect size found by Angrist and Lavy (1999) for Israel using quasi-experimental methods is similar in magnitude, ranging from 0.02 for fourth graders to 0.036 for fifth graders. Our estimates suggest that an increase in the concentration

of recent immigrants in the grade by one standard deviation (5 percentage points) would reduce natives' average test scores by 0.03 standard deviation. Increasing the concentration of recent immigrants in the classroom by one standard deviation, holding class size constant, has therefore a quantitatively similar effect as adding one student to the classroom.

In terms of spending, the STAR project raised costs by about 30% in K-3, and was estimated to raise test scores by 0.17 standard deviation (Krueger and Whitmore (2001)). Spending per pupil in at the time was around \$9,000, so comparable proportional class size reductions would cost around \$2,700 per pupil per year, resulting in a positive effect size of around 0.06 for \$1,000, assuming linear effects of school spending on test scores. Increasing the share of recent migrants in the classroom by one standard deviation (5 percentage points) therefore has a comparable effect as decreasing expenditure per student by \$500 a year.

It is also useful to compare the magnitude of the effects estimated to other studies on peer effects in the classroom. Our estimates of the effect of recent migrants are considerably smaller than that of peer socio-economic background (proxied by books at home) reported by Ammermueller and Pischke (2009) in the European context. Their estimates are drawn from a sample and context similar to ours, using a sample of third graders in European primary schools. They report large effects compared to the US literature and find that a one standard deviation change in their peer composition variable - the average number of books in the home of classroom peers - increases reading test scores by 0.17 of a standard deviation. Other studies on school peer effects typically report size effects ranging from 0.03 to 0.10 of a standard deviation (Lavy et al. (2012a), Duflo et al. (2011)). Our estimated effect of the share of recent migrant in the grade is therefore at the lower end of that range.²²

²²The effect size is similar to that of Lavy et al. (2012a), who find that increasing the share of

One could be concerned that our estimates may be biased towards zero due to missing peer data, an issue highlighted by Ammermueller and Pischke (2009) and Sojourner (2013) in peer effect estimation. Contrary to more detailed parental information collected through the parental questionnaire, however, information on child migration status and duration of stay in the Netherlands was collected from the school administration at the beginning of the school year. As a result, the variable on years spent in the Netherlands used to identify immigrant status and construct our treatment variable - the share of immigrants in the grade - is missing for less than 4% of students in the sample.

To illustrate the limited consequences of missing peer data in our sample, we use the method suggested by Ammermueller and Pischke (2009) which corrects peer effect estimates for missing peer data. As commonly done in the peer effect literature, our treatment variable was constructed using all students in the grade for which immigrant status is not missing. This estimator is referred to as Individual Deletion Procedure (IDP) by Ammermueller and Pischke (2009), who propose a correction method to the IDP estimator which tends to be biased towards zero. Approximately, their correction scales up the IDP point estimate and standard errors by the inverse of the fraction of data observed.²³ Given that peer immigrant status is observed for 96% of observations in our sample, our point estimate adjusted for missing data is -0.032, very similar to our IDP estimate reported in Table 4 (-0.031). The effect of missing peer values on the magnitude of our estimates is therefore minimal.

The evidence reported in section 5.3 suggests that an upward bias in our estimates due to the repeaters in the grade (proxying for the share of low-ability students) by one standard deviation decreases the mean test score of other students by 0.03 of a standard deviation.

²³See Ammermueller and Pischke (2009) for more details on the correction method using the IDP estimator.

allocation of extra school resources to grades with more migrants is unlikely. We can show that, even if our estimates were biased by the non-random allocation of resources to grades with more migrants, the magnitude of the upward bias would be small. Over our period our analysis, the average expenditure per student in primary schools in the Netherlands in the median year of our sample (1999) was about \$4,000 (OECD (2002)). According to the Dutch weighting system, an increase in the share of recent immigrants by one standard deviation (5 percentage points) would lead to an increase of \$180 per student in the grade. As estimates from Krueger and Whitmore (2001) suggest that an additional \$1,000 per student increases test scores by about 0.06 of a standard deviation, the additional amount of resources brought by additional immigrants in the grade would increase test scores by about $0.06 * 180 / 1000 = 0.011$ standard deviation. Accounting for this upward bias would increase the magnitude of our estimate for a one standard deviation increase in the share of recent migrants from -0.031 (Table 4) to -0.042.

6.6 Potential Mechanisms at Play

Previous contributions that report a negative and significant effect of immigrant concentration in the classroom on natives' achievement have not attempted to uncover the mechanisms underlying those effects. The broader literature on peer effects in the classroom, however, suggests at least three main mechanisms through which classroom composition may affect students' achievement. These mechanisms can be adapted to the context of immigrant peer effects to formulate hypotheses about mechanisms at play. First, immigrant students may be more prone to disruption and misbehavior in the classroom relative to native students, therefore reducing teaching time to the entire classroom (disruption hypothesis; Lazear (2001); Lavy and Schlosser (2011)). Second, if achieve-

ment among immigrant students is lower than that of natives, teachers with immigrant students in their classes may provide individualized assistance to those students, which could take instruction time away from other students (crowding-out hypothesis; Lavy et al. (2012a)). Third, teachers may target their instruction to a lower level to accommodate immigrant students if their share is large enough (instruction targeting hypothesis; Duflo et al. (2011), Hunt (2017)). This hypothesis also relies on the assumption that immigrant students' achievement is lower than that of natives'.

While the first hypothesis refers to direct peer effects where individual outcomes are directly affected by the behavior of their peers, the last two can be categorized as indirect peer effects, where spillovers are mediated by teacher behavior in response to classroom composition. Chin et al. (2013) also suggest a fourth possible mechanism. A larger share of immigrant student in the grade may increase administration costs due to the special needs of immigrant students, which may decrease funding allocated to native students. This channel is however expected to be largely muted in the Dutch context, which allocates school resources proportionally to the share of immigrant students in the school.

Despite rich teacher and student questionnaires, our dataset does not include questions that would allow to formally and directly test for these mechanisms, although some questions can be used as corroborative evidence. We therefore restrict ourselves to assess the consistency of the patterns of our results with possible channels, and back it up with additional corroborative evidence when available.

First, the asymmetric spillovers of recent migrants on natives' achievement in language and mathematics appear incompatible with an overall disruption of the classroom through a higher propensity to misbehave and disrupt the class - in the strict sense - among immigrant students

(disruption hypothesis).²⁴ If recent migrants were more prone to disrupt the classroom, resulting in an overall decrease in instruction time and quality for native students, one would expect this to manifest in both subjects, and therefore to observe negative effects on natives tests score irrespective of the subject taught. Patterns of the results also appear incompatible with immigrant students increasing administrative costs and diverting school resources away from natives, as one would also expect this effect to manifest in both subjects. This is unsurprising given the extra funding received by schools based on the share of immigrant students in the school.

The main patterns of our results appear compatible with the crowding out hypothesis and the instruction targeting hypothesis. Contributions on classroom peer effects such as Duflo et al. (2011) or Lavy et al. (2012a) showed that teachers tend to adjust their time allocation or level of instruction in response to classroom composition. Lavy et al. (2012a) find evidence that teachers alter their instruction practices and allocate individualized time to low achievers in the classroom while reducing attention to other students. Duflo et al. (2011) find that tracking students according to prior level of achievement into classrooms has a large positive effect on test scores, and that shifts in teachers instruction to match the achievement composition of the classroom is a key mechanism.

The mechanisms evidenced by Lavy et al. (2012a) and Duflo et al. (2011) indicate that if the achievement composition of the classroom differs in mathematics and Dutch language, one could expect teachers' instruction and time allocation to be affected differently in the two subjects. Particularly, if immigrant students are low achievers in Dutch language relative to mathematics, the effect on teachers' instruction or time allocation could be stronger in Dutch language classes.

²⁴Lazear (2001) however considers a broad definition of disruption in the classroom, which is defined as the proportion of time a given student halts the public aspects of the classroom education process. According to this definition, this would include disruption due to misbehavior (disruption hypothesis), but also times when a student requires teachers' special attention, or times when a student may ask a question that may be known by all classmates (crowding-out hypothesis).

Recent immigrant students could therefore exert more disruption - in the broad sense of Lazear (2001) - in Dutch language classes relative to arithmetic classes, due to greater difficulties in the former subject. While primary teachers teach Dutch language and arithmetic to the same group of students, the two subjects are taught in separate sessions, where they can use separate learning methods and set separate learning objectives for each of the subjects. Primary school teachers also have significant autonomy in choosing what they teach within each subject and how they teach it, in the absence of a centralized decision making regarding curriculum.

The achievement gap between recent immigrant students and natives is considerably larger in Dutch language than it is in mathematics in the PRIMA sample. The average standardized score of recent migrants in Dutch language is about 0.90 standard deviation lower than that of natives, while it is 0.48 standard deviation lower in mathematics. Similarly, the achievement gap of recent migrants relative to natives in Dutch language (0.90 of a standard deviation) is considerably larger than that of migrant students who have been in the country for a longer period relative to natives, in the same subject (0.47 of a standard deviation).

In addition, descriptive evidence on teachers' behaviors in the PRIMA sample suggests that teachers tend to allocate dedicated instruction time to weaker students when they are in the classroom. Among the main teachers, 78% report repeating the basic material several times for students who are weaker in the subject. AND 88% of teachers also report that the extra help provided to students with learning difficulties is often or always given by themselves, as opposed to a remedial teacher or assistant. This descriptive evidence is consistent with the findings of Lavy et al. (2012a) in Israel, and suggests that a higher share of low-achievers in a subject increases instruction time dedicated to weaker students, relative to other students in the classroom (crowding-out hypothesis). According to this mechanism, a larger immigrant-native achievement gap in Dutch language

is compatible with stronger negative spillovers in that subject. It is also compatible with larger effects coming from recent migrants, particularly from those with low prior exposure to Dutch.

Specific patterns of the results are harder to reconcile with the instruction-targeting hypothesis, while they are compatible with the crowding-out hypothesis. If teachers' instruction level was lower for all students in the classroom in the presence of recent immigrant students, e.g. by not covering the more advanced material, one would expect effects to show for all native students in the classroom. The results reported in Table 7 however indicate that negative effects on language score are significant only for natives with low-parental education. This finding can be reconciled with the crowding out hypothesis if teachers spend less individualized time with disadvantaged native students when immigrant students are also present in the classroom, compared to what they would in the absence of immigrant students (crowding-out hypothesis). Another data pattern that makes the instruction-targeting hypothesis less likely in our setting is the fact that recent immigrants are in small numbers in the classroom, typically one or two. Studies that find instruction targeting effects such as that of Duflo et al. (2011) rely on large variation in mean achievement across classrooms, which prompts a shift in teachers instruction level. Although this hypothesis cannot be fully ruled out in our context, the typically small number of recent immigrants in the classroom in our sample makes a shift in teachers instruction level relatively unlikely.

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 8 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.

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 8 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. We also control for school linear trends that could be associated with changes in immigrant concentration within schools.

Results are displayed in Table 9. Our key findings are robust to restricting the sample to one 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 affects 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.

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 10. 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.

8 Conclusion

Our findings contribute to the literature on immigrant peer effects in the classroom by showing that the magnitude of spillover effects depends on the duration of stay of first-generation immigrant classmates in the country. Our results in the Dutch context suggest that immigrant students who have been living in the country for already a few years have no impact on natives' test scores, in either mathematics or language. However, we find that immigrant students who have been in the country for a short period have a small negative effect on natives' performance in language, but no effect on mathematics test scores.

Although the exact mechanisms behind these findings need to be further investigated, the pattern of the results combined with descriptive evidence suggest that assimilation and host country language acquisition play a role in the magnitude of the peer effects generated. 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 negative learning spillovers decline with the duration of stay of immigrants in the host country. Although the evidence we provide to support potential mechanisms behind these findings is descriptive, the pattern of our results suggests that individualized time spent by teachers to support recent immigrant students is the most plausible channel.

These results must however be nuanced by the small size of the estimated effects. An increase by one standard deviation in the share of recent migrants in the classroom is found to reduce natives' average language test scores by 0.03 of a standard deviation. Among studies that estimate the effects of educational interventions and classroom peer effects, our effect size is at the lower end of available estimates. The specificities of the Dutch primary school system, combined with the

balancing tests we conduct to assess the validity of our identification as well as several robustness tests, provide comfort regarding the consistency of our estimates.

Despite the small magnitude of the effects estimated, the key result of this paper highlights the importance of the integration process in offsetting potential adverse effect of immigrant concentration in the classroom. That negative spillovers are small and short-lived suggests that putting in place integration programs for recently arrived migrant students, either in schools or in the host society more broadly, could be sufficient to offset those effects, with a particular focus on host language acquisition. This could be of particular help in schools where native and immigrant children disproportionately 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

Alicia Adsera and Mariola Pytlikova. The Role of Language in Shaping International Migration.

Economic Journal, 125(586):F49–F81, 2015.

Andreas Ammermueller and Jarn-Steffen Pischke. Peer Effects in European Primary Schools: Evidence from the Progress in International Reading Literacy Study. *Journal of Labor Economics*,

27(3):315–348, July 2009.

Joshua D. Angrist and Victor Lavy. Using Maimonides' Rule to Estimate the Effect of Class Size

on Scholastic Achievement. *The Quarterly Journal of Economics*, 114(2):533–575, 1999. URL <https://ideas.repec.org/a/oup/qjecon/v114y1999i2p533-575..html>.

Rosario Maria Ballatore, Margherita Fort, and Andrea Ichino. The Tower of Babel in the Classroom: Immigrants and Natives in Italian Schools. CEPR Discussion Papers 10341, C.E.P.R. Discussion Papers, January 2015. URL <https://ideas.repec.org/p/cpr/ceprdp/10341.html>.

Anders Bohlmark. Age at immigration and school performance: A siblings analysis using swedish register data. *Labour Economics*, 15(6):1366–1387, December 2008. URL <https://ideas.repec.org/a/eee/labeco/v15y2008i6p1366-1387.html>.

Maria Branden, Elisabeth Birkelund, and Ryszard Szulkin. Does school segregation lead to poor educational outcomes? evidence from fifteen cohorts of swedish ninth graders. The Institute for Analytical Sociology Working Paper Series, 2016.

Thibault Brodaty. Peer effects in education: A literature review. *Revue d'économie politique*, 120(5):739–757, 2010. doi:10.3917/redp.205.0739.

Giorgio Brunello and Lorenzo Rocco. The effect of immigration on the school performance of natives: Cross country evidence using PISA test scores. *Economics of Education Review*, 32(C):234–246, 2013. doi:10.1016/j.econedurev.2012. URL <https://ideas.repec.org/a/eee/econedu/v32y2013icp234-246.html>.

Scott E. Carrell, Richard L. Fullerton, and James E. West. Does Your Cohort Matter? Measuring Peer Effects in College Achievement. *Journal of Labor Economics*, 27(3):439–464, July 2009. URL <https://ideas.repec.org/a/ucp/jlabec/v27y2009i3p439-464.html>.

Aimee Chin, N. Meltem Daysal, and Scott A. Imberman. Impact of bilingual education programs on limited English proficient students and their peers: Regression discontinuity evidence from Texas. *Journal of Public Economics*, 107(C):63–78, 2013. doi:10.1016/j.jpubeco.2013.08. URL <https://ideas.repec.org/a/eee/pubeco/v107y2013icp63-78.html>.

Kalena E. Cortes. The effects of age at arrival and enclave schools on the academic performance of immigrant children. *Economics of Education Review*, 25(2):121–132, April 2006. URL <https://ideas.repec.org/a/eee/eoedu/v25y2006i2p121-132.html>.

Timothy M. Diette and Ruth Uwaifo Oyelere. Gender and racial differences in peer effects of limited English students: a story of language or ethnicity? *IZA Journal of Migration and Development*, 6(1):1–18, December 2017. doi:10.1186/s40176-017-0111-5. URL https://ideas.repec.org/a/spr/izamig/v6y2017i1d10.1186_s40176-016-0074-y.html.

Simone Dobbelsteen, Jesse Levin, and Hessel Oosterbeek. The Causal Effect of Class Size on Scholastic Achievement: Distinguishing the Pure Class Size Effect from the Effect of Changes in Class Composition. *Oxford Bulletin of Economics and Statistics*, 64(1):17–38, February 2002. URL <https://ideas.repec.org/a/bla/obuest/v64y2002i1p17-38.html>.

Esther Duflo, Pascaline Dupas, and Michael Kremer. Peer Effects, Teacher Incentives, and the Impact of Tracking: Evidence from a Randomized Evaluation in Kenya. *American Economic Review*, 101(5):1739–1774, August 2011. URL <https://ideas.repec.org/a/aea/aecrev/v101y2011i5p1739-74.html>.

D. Epple and R. Romano. Peer effects in education: Survey of the theory and evidence. In *Hand-*

book of Social Economics, volume 1, pages 1053–1063. Elsevier BV, 2011. doi:10.1016/B978-0-444-53707-2.00003-7.

Tommaso Frattini and Elena Meschi. The Effect of Immigrant Peers in Vocational Schools. IZA Discussion Papers 11027, Institute for the Study of Labor (IZA), September 2017. URL <https://ideas.repec.org/p/iza/izadps/dp11027.html>.

Charlotte Geay, Sandra McNally, and Shqiponja Telhaj. Non-native Speakers of English in the Classroom: What Are the Effects on Pupil Performance? *Economic Journal*, 0:281–307, August 2013. URL <https://ideas.repec.org/a/ecj/econjl/vy2013ipf281-f307.html>.

Eric D. Gould, Victor Lavy, and M. Daniele Paserman. Does Immigration Affect the Long-Term Educational Outcomes of Natives? Quasi-Experimental Evidence. *Economic Journal*, 119(540):1243–1269, October 2009. URL <https://ideas.repec.org/a/ecj/econjl/v119y2009i540p1243-1269.html>.

Eric A. Hanushek. *School Resources*, volume 2 of *Handbook of the Economics of Education*. Elsevier, June 2006.

Eric A. Hanushek, John F. Kain, Jacob M. Markman, and Steven G. Rivkin. Does peer ability affect student achievement? *Journal of Applied Econometrics*, 18(5): 527–544, 2003. doi:10.1002/jae.740. URL <https://ideas.repec.org/a/jae/japmet/v18y2003i5p527-544.html>.

Eric A. Hanushek, John F. Kain, and Steven 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, July 2009. URL <https://ideas.repec.org/a/ucp/jlabec/v27y2009i3p349-383.html>.
- James Heckman and Flavio Cunha. The Technology of Skill Formation. *American Economic Review*, 97(2):31–47, May 2007. URL <https://ideas.repec.org/a/aea/aecrev/v97y2007i2p31-47.html>.
- Jennifer Hunt. The Impact of Immigration on the Educational Attainment of Natives. *Journal of Human Resources*, 52(4):1060–1118, 2017. URL <https://ideas.repec.org/a/uwp/jhriss/v52y2017i4p1060-1118.html>.
- Peter Jensen and Astrid Wrtz 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. URL <https://ideas.repec.org/a/eee/eoedu/v30y2011i6p1503-1515.html>.
- Alan B. Krueger. Experimental Estimates of Education Production Functions. *The Quarterly Journal of Economics*, 114(2):497–532, 1999.
- Alan B. Krueger and Diane M. Whitmore. The Effect of Attending a Small Class in the Early Grades on College-Test Taking and Middle School Test Results: Evidence from Project STAR. *Economic Journal*, 111(468):1–28, January 2001. URL <https://ideas.repec.org/a/ecj/econjl/v111y2001i468p1-28.html>.
- Helen F. Ladd and Edward B. Fiske. Weighted student funding in the Netherlands: A model for the U.S.? *Journal of Policy Analysis and Management*, 30(3):470–498, June 2011.

Victor Lavy and Analia Schlosser. Mechanisms and Impacts of Gender Peer Effects at School. *American Economic Journal: Applied Economics*, 3(2):1–33, April 2011. URL <https://ideas.repec.org/a/aea/aejapp/v3y2011i2p1-33.html>.

Victor Lavy, M. Daniele Paserman, and Analia Schlosser. Inside the Black Box of Ability Peer Effects: Evidence from Variation in the Proportion of Low Achievers in the Classroom. *Economic Journal*, 122(559):208–237, March 2012a. doi:j.1468-0297.2011.02462.x. URL <https://ideas.repec.org/a/ecj/econjl/v122y2012i559p208-237.html>.

Victor Lavy, Olmo Silva, and Felix Weinhardt. The Good, the Bad, and the Average: Evidence on Ability Peer Effects in Schools. *Journal of Labor Economics*, 30(2):367–414, 2012b. URL <https://ideas.repec.org/a/ucp/jlabec/doi10.1086-663592.html>.

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

Charles F. Manski. Identification of Endogenous Social Effects: The Reflection Problem. *Review of Economic Studies*, 60(3):531–542, 1993. URL <https://ideas.repec.org/a/oup/restud/v60y1993i3p531-542..html>.

Florence Neymotin. Immigration and its effect on the college-going outcomes of natives. *Economics of Education Review*, 28(5):538–550, October 2009.

OECD. Education at a glance 2002. Paris: Organisation for Economic Co-Operation and Development, 2002.

Asako Ohinata and Jan 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.

Asako Ohinata and Jan C. van Ours. How Immigrant Children Affect the Academic Achievement of Native Dutch Children. *Economic Journal*, 0:308–331, August 2013. URL <https://ideas.repec.org/a/ecj/econjl/vy2013ipf308-f331.html>.

Asako Ohinata and Jan C. van Ours. Quantile Peer Effects of Immigrant Children at Primary Schools. *LABOUR*, 30(2):135–157, June 2016. URL <https://ideas.repec.org/a/bla/labour/v30y2016i2p135-157.html>.

Bruce Sacerdote. Peer Effects with Random Assignment: Results for Dartmouth Roommates. *The Quarterly Journal of Economics*, 116(2):681–704, 2001. URL <https://ideas.repec.org/a/oup/qjecon/v116y2001i2p681-704..html>.

Nicole Schneeweis. Immigrant concentration in schools: Consequences for native and migrant students. *Labour Economics*, 35(C):63–76, 2015. doi:10.1016/j.labeco.2015.05. URL <https://ideas.repec.org/a/eee/labeco/v35y2015icp63-76.html>.

Aaron Sojourner. Identification of Peer Effects with Missing Peer Data: Evidence from Project STAR. *Economic Journal*, 123(569):574–605, June 2013. URL <https://ideas.repec.org/a/ecj/econjl/v123y2013i569p574-605.html>.

Marco Tonello. Peer effects of non-native students on natives educational outcomes: mechanisms and evidence. *Empirical Economics*, 51(1):383–414, August 2016. doi:10.1007/s00181-015-0994-z. URL https://ideas.repec.org/a/spr/empeco/v51y2016i1d10.1007_s00181-015-0995-y.html.

Jan C. van Ours and Justus 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

	Native Dutch	Immigrants			
		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 (10.11)	41.21 (10.37)	40.18 (10.13)	42.55 (10.23)	42.04 (10.83)
Father's education: lower secondary	49.51 (9.43)	44.52 (10.57)	41.09 (10.11)	44.18 (10.24)	44.35 (10.50)
Father's education: upper secondary	52.66 (8.67)	46.89 (10.54)	42.96 (10.46)	45.36 (10.33)	45.97 (10.48)
Father's education: university	55.10 (7.99)	48.84 (10.53)	46.07 (9.34)	48.99 (10.09)	47.20 (10.37)
All students	50.46 (9.79)	44.94 (10.89)	40.85 (10.19)	44.23 (10.36)	43.05 (10.58)
Average test score – mathematics					
Father's education: primary	45.74 (10.36)	44.70 (10.84)	44.34 (10.53)	43.43 (10.62)	45.26 (10.79)
Father's education: lower secondary	49.12 (9.86)	46.26 (10.53)	45.21 (10.21)	44.11 (10.71)	46.65 (11.13)
Father's education: upper secondary	52.05 (9.01)	48.35 (10.16)	47.41 (9.77)	45.34 (10.23)	47.96 (10.58)
Father's education: university	54.34 (8.31)	50.50 (10.22)	50.05 (10.05)	47.69 (10.11)	49.71 (10.16)
All students	50.29 (9.88)	46.69 (10.76)	45.01 (10.44)	44.21 (10.38)	46.89 (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 (13.02)	22.03 (12.36)	29.79 (14.11)	23.11 (11.23)	21.96 (11.49)	29.24 (15.82)
Fraction of immigrant students	0.063 (0.133)	-	0.053 (0.021)	0.137 (0.027)	0.279 (0.072)	0.820 (0.180)
Share of natives with low parental education	0.446 (0.497)	0.406 (0.491)	0.441 (0.497)	0.560 (0.496)	0.598 (0.490)	0.582 (0.493)
Average test score in Dutch language						
All students	49.89 (5.37)	51.13 (5.01)	49.77 (4.91)	47.37 (5.46)	45.72 (5.41)	45.93 (5.65)
Immigrant students	44.93 (9.13)	-	45.73 (10.12)	44.10 (8.21)	43.24 (6.86)	45.57 (6.09)
Native students	50.14 (5.42)	51.22 (5.04)	50.81 (4.97)	47.89 (5.71)	46.63 (5.79)	44.87 (8.15)
Average test score in mathematics						
All students	49.88 (4.94)	50.76 (4.82)	49.76 (4.58)	48.08 (4.95)	46.99 (5.12)	47.41 (4.76)
Immigrant students	46.86 (9.03)	-	47.30 (10.05)	46.37 (8.10)	45.94 (7.10)	47.12 (5.07)
Natives	50.00 (5.04)	50.81 (4.87)	49.89 (4.68)	48.35 (5.16)	47.34 (5.33)	46.99 (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. Native students with low parental education are defined as students whose father did not complete upper secondary education.

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

Dep. variable: % of immigrants in the grade	OLS (1)	School F.E. (2)	School-by-year F.E. (3)	School-by-year F.E.+linear trend (4)
% of natives whose father has primary education	0.147*** (0.013)	-0.010 (0.007)	-0.001 (0.006)	-0.004 (0.005)
% of natives whose father has lower secondary education	-0.012 (0.012)	0.002 (0.005)	0.001 (0.004)	0.002 (0.004)
% of natives whose father has upper secondary education	-0.132*** (0.013)	-0.012* (0.006)	-0.001 (0.006)	-0.001 (0.004)
% of natives whose father has tertiary education	-0.092*** (0.012)	0.014 (0.008)	-0.004 (0.006)	0.001 (0.005)
Fraction of female students	0.04** (0.014)	0.014 (0.009)	-0.002 (0.005)	0.003 (0.005)
Fraction of natives from disadvantaged families	0.105*** (0.010)	0.010 (0.008)	0.006 (0.006)	0.002 (0.004)
Average class size	-0.092*** (0.010)	-0.007 (0.010)	-0.010 (0.010)	-0.002 (0.008)
Total number of students in the grade	0.038*** (0.009)	-0.018** (0.008)	-0.013 (0.017)	-0.002 (0.007)
Fraction of natives that repeated a grade	0.102*** (0.014)	-0.012 (0.009)	-0.007 (0.006)	0.001 (0.004)
Teacher's years of experience	0.067*** (0.011)	-0.013 (0.010)	-0.006 (0.008)	0.001 (0.007)
Whether the grade has a remedial teacher	0.140*** (0.052)	0.237** (0.095)	0.089 (0.081)	0.102 (0.78)
Whether the class is split for part of the instruction	0.028 (0.044)	0.039 (0.044)	0.041 (0.044)	0.039 (0.043)
Whether the grade has a teaching assistant	0.225* (0.115)	-0.021 (0.251)	-0.021 (0.251)	-0.025 (0.230)
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 share of immigrant students in the grade on the corresponding explanatory variable. Robust standard errors clustered at the school level are in parentheses. Regressions include grade dummies. The last row was estimated for 2004 only as the information was collected only in that year.

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 (0.697)	-0.751 (0.769)	-0.185 (0.701)	-0.191 (0.706)
Treatment 2: Share of recent immigrants	-3.08*** (0.981)	-2.88*** (0.977)	-1.55 (0.913)	-1.49 (0.906)
Treatment 3: Share of other immigrants	0.145 (0.688)	0.168 (0.676)	0.012 (0.691)	-0.027 (0.653)
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, the total number of students in the grade and its square. Controls for linear school trends are also included. Estimates that do not control for linear school trends are quantitatively similar and available upon request.

Table 5: Effects of the share of immigrant classmates by country group of origin

	Natives'		Natives'	
	language score		math score	
	(1)	(2)	(3)	(4)
Treatment 4: Share of immigrants of Turkish/Moroccan origin	-1.73 (1.782)	-1.65 (1.701)	-0.200 (2.287)	-0.222 (1.630)
Treatment 5: Share of immigrants of origin other than Turkish/Moroccan	-0.033 (0.863)	0.588 (0.818)	0.735 (1.104)	0.846 (0.884)
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, the total number of students in the grade and its square. Controls for linear school trends are also included. Estimates that do not control for linear school trends are quantitatively similar and available upon request.

Table 6: Effects of the share of immigrant classmates with low prior exposure to Dutch

	Natives' language score		Natives' math score	
	(1)	(2)	(3)	(4)
Treatment 6: Share of recent immigrants with low prior exposure to Dutch	-3.66** (1.470)	-3.68** (1.450)	-0.872 (1.271)	-0.856 (1.226)
Treatment 7: Share of other immigrants with low prior exposure to Dutch	-1.56 (1.375)	-1.48 (1.362)	-0.608 (1.206)	-0.618 (1.290)
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, the total number of students in the grade and its square. Controls for linear school trends are also included. Estimates that do not control for linear school trends are quantitatively similar and available upon request.

Table 7: 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 immigrants in grade	-1.131 (1.291)	-1.184 (1.223)	-0.935 (1.178)	-0.620 (1.171)	-3.34** (1.457)	-3.35** (1.446)	-1.13 (1.402)	-1.12 (1.379)
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 immigrants in grade	0.202 (1.425)	0.156 (1.375)	0.090 (1.502)	0.118 (1.563)	-1.503 (1.657)	-1.462 (1.276)	-0.608 (1.643)	0.587 (1.576)
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, the total number of students in the grade and its square. 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. Controls for linear school trends are also included. Estimates that do not control for linear school trends are quantitatively similar and available upon request.

Table 8: 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*** (0.981)	-2.88*** (0.977)	-1.55 (0.913)	-1.49 (0.906)
Placebo 1: Share of recent immigrants in next grade	-0.361 (1.241)	0.120 (0.112)	0.038 (1.201)	0.179 (1.192)
Placebo 2: Share of recent immigrants in previous grade	0.526 (1.189)	0.534 (1.181)	-0.472 (1.092)	-0.294 (1.071)
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, the total number of students in the grade and its square. Controls for linear school trends are also included. Estimates that do not control for linear school trends are quantitatively very similar and available upon request.

Table 9: 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 (1.336)	-1.200 (1.273)	-0.985 (1.400)	-0.957 (1.376)
Treatment 2: Share of recent immigrants	-3.78** (1.301)	-3.72** (1.210)	-1.75 (1.415)	-1.67 (1.302)
Treatment 3: Share of other immigrants	0.145 (1.347)	0.168 (1.267)	0.142 (1.368)	0.128 (1.333)
Grade-level controls		✓		✓
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, the total number of students in the grade and its square. Controls for school linear trends are also included. Estimates that do not control for linear school trends are quantitatively very similar and available upon request.

Table 10: 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*** (0.981)	-2.88*** (0.977)	-1.55 (0.913)	-1.49 (0.906)	-3.60*** (1.28)	-3.34*** (1.38)	-1.48 (1.27)	-0.980 (1.19)
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, the total number of students in the grade and its square. Controls for linear school trends are also included. Estimates that do not control for linear school trends are quantitatively very similar and available upon request.

Appendix

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

	OLS	School F.E.	School-by-year F.E.	School-by-year F.E.+linear trend
Dependent variable: % of recent immigrants in grade				
% of natives whose father has primary education	0.162*** (0.014)	-0.009 (0.007)	-0.002 (0.008)	-0.002 (0.007)
% of natives whose father has lower secondary education	-0.011 (0.0122)	0.003 (0.005)	-0.001 (0.006)	0.001 (0.006)
% of natives whose father has upper secondary education	-0.145*** (0.013)	-0.015* (0.007)	0.001 (0.007)	-0.001 (0.005)
% of natives whose father has university education	-0.113*** (0.012)	0.017* (0.008)	-0.005 (0.008)	-0.001 (0.006)
Fraction of female students	0.051** (0.014)	0.014 (0.009)	-0.002 (0.006)	-0.003 (0.005)
Fraction of natives from disadvantaged families	0.102*** (0.013)	0.012 (0.008)	0.003 (0.005)	0.001 (0.004)
Average class size	-0.071*** (0.012)	-0.012 (0.010)	-0.010 (0.011)	-0.002 (0.008)
Total number of students in the grade	0.041*** (0.011)	-0.020** (0.008)	-0.011 (0.016)	-0.001 (0.007)
Fraction of students that repeated a grade	0.131*** (0.014)	-0.014 (0.009)	-0.005 (0.008)	-0.002 (0.007)
Teacher's years of experience	0.098*** (0.019)	0.011 (0.012)	0.003 (0.007)	0.001 (0.006)
Whether the grade has a remedial teacher	0.668*** (0.052)	0.820*** (0.170)	0.030 (0.132)	0.051 (0.124)
Whether the class is split for part of the instruction	0.070 (0.173)	0.067 (0.162)	0.058 (0.164)	0.062 (0.160)
Whether the grade has a teaching assistant	1.12*** (0.499)	1.01 (0.659)	1.01 (0.659)	1.10 (0.643)
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 share of immigrant students in the grade on the corresponding explanatory variable. Robust standard errors clustered at the school level are in parentheses. Regressions include grade dummies. The last row was estimated for 2004 only as the information was collected only in that year.