Linguistic diversity in the classroom, student achievement, and social integration

Julia Bredtmanna,b,c, Sebastian Ottena,c,d and Christina Vonnahmea,e

aRWI – Leibniz Institute for Economic Research, Essen, Germany; bIZA, Bonn, Germany; cCReAM, London, UK; dUniversity of Duisburg-Essen, Essen, Germany; eRuhr University Bochum, Bochum, Germany

ABSTRACT
We analyze whether non-native speakers in the classroom affect students’ educational achievement and social integration. In contrast to previous studies, which mainly examine the effect of the share of immigrant pupils, we focus on language heterogeneity by using a novel measure of the degree of linguistic diversity in the classroom. Conditional on the concentration of non-native speakers in the class, the degree of linguistic diversity has no adverse effect on students’ language and math skills, but worsens the social integration of non-native speakers. We demonstrate the robustness of these findings in a variety of sensitivity checks.

ARTICLE HISTORY
Received 25 April 2019
Accepted 10 December 2020

KEYWORDS
Immigration; linguistic diversity; educational performance; social integration

JEL CLASSIFICATIONS
I21; I24; J15

1. Introduction
Over the last decades, many Western countries experienced large inflows of immigrants. For instance, Germany witnessed a huge increase in immigration due to both rising immigrant populations from within the EU and a high number of refugees. In 2015, the net immigration of foreign people to Germany reached 1.24 million, which represents an 84% increase compared to 2014 and a record high in post-war history (BAMF 2015). The increase in immigration has drawn considerable attention to issues regarding the impact of immigrants on labor market outcomes of natives (e.g. Card 1990, 2001; Borjas 2003; Dustmann, Frattini, and Preston 2013; Foged and Peri 2016; Dustmann, Schönberg and Stuhler 2017) and the fiscal effects of immigration (e.g. Auerbach and Oreopoulos 1999; Dustmann and Frattini 2014; Preston 2014). Due to the rising share of immigrant students in destination countries, similar debates about the integration of immigrant children have opened in recent years. As children of immigrants exhibit significant gaps in school performance relative to native children (Lüdemann and Schwerdt 2013; Giannelli and Rapallini 2016; Ruhose and Schwerdt 2016), concerns about adverse effects of immigrant children on the educational outcomes of native children have been raised.1 These concerns might reinforce the tendency of native parents to choose schools with a low immigrant concentration, thus fostering ethnic segregation in schools.

Despite being a key part of the immigration debate, the literature on the role of immigrant children in schools is relative small and the evidence remains largely inconclusive. Previous research has mainly focused on analyzing the effects of immigrant concentration in classes or schools on the educational outcomes of both immigrant and native children. However, rising migration
flows do not only increase the share of immigrant children in schools, but the fact that current immigrants to Europe increasingly come from more culturally and linguistically distant countries also changes the ethnic and linguistic composition of the class. Yet, relatively little evidence exists on how the degree of ethnic or linguistic diversity in the classroom affects student outcomes.

In this paper, we examine whether – in addition to the concentration of non-native speakers in the class – the linguistic composition of this group matters for native and migrant students’ educational outcomes. In particular, we analyze whether the degree of linguistic diversity in the class has an impact on the language and math test scores as well as on the social integration of native and non-native speakers. We thus go beyond studies that merely analyze the effect of a certain share of immigrant students in a class on school performance. Analyzing the role of linguistic diversity has important implications for the optimal allocation of immigrant students to classes and is thus of utmost interest for both policymakers and educators, as linguistic diversity can typically be influenced more easily compared to the share of immigrant students.

To provide conclusive evidence on this issue, we rely on contributions from the macroeconomic and political science literature (e.g. Easterly and Levine 1997; Alesina et al. 2003; Montalvo and Reynal-Querol 2005) and construct a novel measure of the degree of linguistic diversity in the class, which takes into account both the size of the different immigrant groups and the linguistic distance between them. To address the problem of a potential endogeneity of the degree of linguistic diversity, we analyze the effect of linguistic diversity conditional on the share of non-native speakers in the class, thus comparing classes that are observationally equivalent in terms of the sorting of native and non-native speakers.

Our analysis is based on a comprehensive survey of 4th-grade students in German primary schools. The dataset has the rare feature of containing detailed information on students’ and their parents’ migration history, on children’s mother tongue, family and school characteristics, as well as results of standardized tests in both German language and math. Information on students’ social integration in the class further allows us to shed light on whether social cohesion is affected by the linguistic diversity in the class.

Our results reveal a negative association between the share of non-native speakers in the class and students’ test scores and their social integration. Conditional on the concentration of non-German speakers in the class, the degree of linguistic diversity has no impact on students’ language and math test scores. This reveals that an increase in the number of students from more culturally or linguistically distant countries has no additional negative impact on students’ educational outcomes. We find, however, that a higher linguistic diversity in the class hampers the social integration of non-native speakers. In particular, non-native students in classes with a high linguistic diversity are more likely to have arguments with their classmates and have less friends in class. This suggests that the social integration of migrants could be improved by reducing linguistic diversity and allocating more students with the same linguistic background to the same class.

Previous literature has mainly assessed the effects of class composition by analyzing how the share of migrants in the class or school affects the educational outcomes of native and non-native students. The results of these studies are mixed. Most studies find an adverse effect of immigrant peers on the school performance of non-native students (Jensen and Rasmussen 2011; Ohinata and van Ours 2013; Schneeweis 2015). However, evidence for native students is inconclusive. While Gould, Lavy, and Paserman (2009), Jensen and Rasmussen (2011), Cho (2012), Diette and Oyelere (2014, 2017), Tonello (2016) and Ballatore, Fort, and Ichino (2018) find negative effects, Ohinata and van Ours (2013) and Schneeweis (2015) find no effect of immigrant concentration on the school performance of native pupils.

While several studies analyze the consequences of immigrant concentration in schools for students’ educational outcomes, much less is known about the effects of the composition of the immigrant group in a class. However, when investigating the effects of immigration, it is reasonable to argue that it is not only the share of migrants or non-native speakers that matters for students’
outcomes, but that the degree of diversity among them is also relevant. As language proficiency is a strong predictor of children’s schooling success, especially the degree of linguistic diversity in the classroom should matter for students’ outcomes. On the one hand, the grouping of children with a similar mother tongue may improve their self-consciousness through identity-building and thus foster their learning outcomes. On the other hand, the formation of a large group of children with a different mother tongue than the majority language may slow the learning of this language and negatively impact learning by dividing the class and impeding the children’s sense of togetherness.

The few existing studies that analyze the effects of class- or school-level diversity on students’ outcomes focus on ethnic diversity. In particular, these studies use variants of the Herfindahl-Hirschman index calculated based on students’ ethnicity or their (parents’) country of birth to measure ethnic diversity within the class or school. Using PISA data of 15-year-old students from 15 OECD countries, Dronkers and van der Velden (2013) find a negative association between the ethnic diversity in the school and the language performance of immigrant students. The language performance of native students is only negatively influenced in highly stratified educational systems. The study by Maestri (2017) uses data of Dutch primary school students and finds that ethnic diversity has no impact on native students’ literacy scores, but does increase those of immigrant students. The results further suggest a negative effect of ethnic diversity on social integration. Frattini and Meschi (2019) use administrative data on the universe of students in Italian vocational training institutions. They find that the presence of immigrant students in the classroom has no effect on native students’ literacy achievements but small negative effects on their math scores. Ethnic diversity, however, has no effect on students’ performance.

Our work contributes to the literature on the externalities of non-native peers on students’ educational outcomes in several dimensions. First, we add to the small literature that analyzes the effect of the composition of the immigrant group on native and non-native students’ school performance. While previous studies focus on ethnic heterogeneity to measure diversity in the class, we instead use information on students’ mother tongue, which is particularly relevant in the context of language acquisition and application at school. Moreover, we extend the previously used diversity measure, the Herfindahl-Hirschman index, by incorporating a component measuring the distance between the different language groups into the diversity measure. The resulting Greenberg index is a more precise measure of the degree of linguistic heterogeneity in the class. In the context of social interactions in education, this is the first time that a diversity measure takes into account both the size of the different immigrant groups and the distance between them.

Second, we extend the empirical literature on Germany, for which evidence on the effects of immigration on students’ educational outcomes is still rare. Germany presents an interesting case study, as it is a country with a long migration history where concerns against immigrants and the question of how to integrate them best in the educational systems have recently become highly topical given high absolute and relative numbers of immigrants, in particular the 400,000 school-age refugees (The Economist 2017; Spiegel Online 2017).

In contrast to the existing literature that focuses mainly on students in high schools and above, we further contribute to the literature by analyzing the effects of immigrant students on primary school children. This allows us to evaluate the effects on native and non-native students at a young age. Knowledge about the effects at a young age is particularly important as the foundation for success in school and in the labor market is already laid in the first years of schooling. The benefits of a high quality early childhood education are especially high for disadvantaged and immigrant children (Arnold and Doctoroff 2003; Heckman 2006). Moreover, the existing evidence on high school students is likely to reflect the accumulated impact from the exposure to immigrant students during many years of schooling. Studying young students has the advantage of reducing the extent of such an accumulated exposure effect.

Lastly, we contribute to the literature by analyzing whether the concentration and the composition of the non-native speakers in the class affect the social integration of students. This is an
important question as good student-student relationships are a key factor in creating a positive classroom climate (Kyriakides and Creemers 2008), which itself can be an important determinant of children’s school success. Furthermore, the integration of immigrants and their children in the society of their destination country is one of the major challenges immigration countries face. Schools can provide the ideal environment to improve integration and reduce the difficulties faced by immigrant children. Providing evidence on the effects of class composition on migrants’ social integration is therefore of particular interest for both policymakers and educators.

The remainder of the paper is as follows. Section 2 describes the data used and the construction of the linguistic diversity measure. In Section 3, we outline the empirical framework. Results and sensitivity analyses are discussed in Section 4. Section 5 provides concluding remarks.

2. Data and diversity measure

We analyze how linguistic diversity in the classroom affects the educational success of schoolchildren in Germany, a country with a high share of children with migration background that has been rising steadily in the past years. For this purpose, we use data from the ‘Ländervergleich 2011’ (which literally means ‘comparison of [federal] states’), a survey with a gross sample of 27,081 4th-grade students in 1349 schools in all German federal states. The dataset is provided by the Research Data Centre (FDZ) at the Institute for Educational Quality Improvement (IQB). The main goal of the survey was to systematically compare the achievements in German and math of children at the end of primary school, when they are typically between 9 and 10 years old. Analyzing primary school students is particularly suitable for our research question, as in the German school system parents’ primary school choice is restricted by legal regulations. Therefore, other than for high-schools, parents cannot freely choose a primary school for their children. Children are assigned to primary schools based on school districts. Thus, all children who reside in a specific school district are obligated to enroll in the primary school of this district to reduce sorting into schools. In addition, at the school level, principals can regulate the allocation of students to classes, and aim at balancing the class composition within a grade, for example with respect to children’s immigration status or socio-economic background.

The sampling procedure of the survey first randomly selected primary schools in each federal state, and then randomly picked one class within each school. The dataset is particularly suitable for our research question as the sample size is large compared to other datasets. Moreover, it has the rare feature of containing detailed information on the students’ and their parents’ migration history, on children’s mother tongue, socioeconomic and school characteristics, and results of standardized tests in both German language (reading and listening) and math (five different learning fields: numbers and operations; space and form; patterns and structures; quantities and measures; data, frequencies and probabilities). The tests and surveys were the same for all students, and all students per class were included. The surveys covered the students, their parents, teachers and schoolmasters.

The information on children’s mother tongue, which is provided by the children’s parents, is key to distinguish between students whose mother tongue is German (‘native speakers’) and students whose mother tongue is a language other than German (‘non-native speakers’), and to compute our measure of linguistic diversity in the classroom. Mother tongue refers to the language the parents used first when speaking with their child, thus it is the language the child learned first. For confidentiality reasons, information on mother tongue is restricted to the ten most prevalent languages in the data. We therefore impute further languages by using information on the countries of birth of the children as well as of their parents and grandparents to increase the sample size and keep it representative for the population of children with migration background. This is done for countries where a specific language can be unambiguously assigned and is spoken by the vast majority of the population.
The resulting distribution of 17 mother tongues is shown in Figure 1. Turkish is the language spoken by most non-native speakers with a share of almost 50%. The high concentration of Turkish-speaking students in German schools is due to the fact that children of Turkish immigrants are significantly less likely to have German as their mother tongue than children of other groups of immigrants. While the share of Turkish students among all immigrant students in our sample amounts to only 22%, the share of Turkish speakers among all non-native speakers is thus much higher. The fact that children of Turkish descent are less likely to speak German at home than other groups of immigrants has already been established in the literature (see, e.g. Casey and Dustmann 2008; Cornelissen et al. 2018). Further important mother tongues of school children in Germany include Russian, Polish and Kurdish with shares of 5 to 17% each (see Figure 1).

Based on the information on children’s mother tongue, we calculate our two main variables of interest, the share of non-native speakers and the linguistic diversity in the class. To construct the latter, we rely on contributions from the macroeconomic and political science literature, where different diversity measures have been refined and applied to analyze the impact of ethnolinguistic diversity on economic growth, redistribution, and measures of political stability (e.g. Easterly and Levine 1997; Alesina et al. 2003; Montalvo and Reynal-Querol 2005). Among the different measures, which can be summarized after Desmet, Ortuño-Ortín, and Weber (2009) as measures of ‘social effective antagonism’, the so-called Greenberg index (Greenberg 1956) stands out to suitably measure differential effects of linguistic diversity in the classroom on student performance. It is defined as

\[ \text{GI} = \sum_{j=1}^{N} \sum_{k=1}^{N} s_{jk} \delta_{jk} \]  

(1)
and increases in the number of (language) groups, \(N\), and the similarity of the relative size of the different groups \(j, s_j\). \(\delta_{jk}\) measures the (linguistic) distance between each pair of groups \(j\) and \(k\). By incorporating the distance between the groups, the diversity measure accounts for the heterogeneity between the groups leading to higher index values for a more distant set of groups. With \(\delta_{jk}\) and the group shares \(s_j\) scaled between zero and one, the Greenberg index ranges between zero and one where larger values indicate higher diversity. It is closely related to the simple measure of (ethno-linguistic) fractionalization, given as

\[
\text{ELF} = 1 - \sum_{j=1}^{N} s_j^2,
\]

which is the reverse of the commonly used Herfindahl-Hirschman index:

\[
\text{HHI} = \sum_{j=1}^{N} s_j^2.
\]

In contrast to these measures without distance, the Greenberg index is not ‘color-blind’, i.e. it not only takes into account the number and size of different groups, but also one further characteristic, which is the linguistic distance to other groups. The color-blindness is described as a deficit of the Herfindahl-Hirschman index amongst others by Dronkers and van der Velden (2013), who use this measure in their analysis. Desmet, Ortuño-Ortín, and Weber (2005) show that for the effect of diversity on redistribution, the linguistic distance between the languages is highly important. We therefore incorporate linguistic distance into our analysis using a new measure of linguistic distance developed by linguists (see Bakker et al. 2009), which has recently been applied to the economic context by Isphording and Otten (2013, 2014). The so-called ‘Levenshtein distance’ is computed by comparing the phonetic similarity of each word of a given word list for each pair of languages. The ‘Swadesh word list’ (Swadesh 1952) includes 40 standard words with translations in all languages. The average distance in the phonetic transcription between two languages is scaled between zero (no linguistic distance) and 100 (maximum linguistic distance). We rescale it to range between zero and one to include it in our measure for linguistic diversity.

As the share of German speakers is already captured by our second variable of interest, the percentage of non-native speakers in the class, we calculate linguistic diversity within the group of non-native speakers in the class. The mean of the resulting measure of linguistic diversity is 0.33 for the sample of non-native speakers and 0.09 for the sample of native speakers, as shown in Table A1 in the Online Appendix. It is, by definition, zero for classes without non-native speakers, which we explicitly capture by including a respective indicator variable for these classes.\(^{10}\)

The share of non-native speakers in the class is 0.28 for the sample of non-native speakers and 0.06 for the sample of native speakers.\(^{11}\) This already points to a certain concentration of non-native speakers in schools, or more generally to a segregation of natives and migrants. Fifty percentage of the children in the sub-sample of native speakers are in classes in which none of the children has a mother tongue other than German. The distribution of the share of non-native speakers for both sub-samples is displayed in Figure A1 in the Online Appendix.

To measure students’ performance in school, we rely on the results of the standardized tests in German language and math that have been conducted as part of the survey. The standardized test scores are scaled to a mean of 500 and a standard deviation of 100 in the gross sample, as is common for several education datasets including the PISA data. In our sample, children whose native language is German on average achieve test scores of 522 and 521 points in language and math, respectively. In the sample of non-native speakers, the scores are substantially lower with 450 points in language and 456 points in math. The standard deviation is slightly higher in the non-native speaker sample. To analyze if linguistic diversity has an impact on
children’s cohesion, i.e. beyond individual performance, we use a measure of social integration as a further dependent variable. The social integration index is based on four questions on children’s relations with classmates. The children indicated whether they agreed: (1) not, (2) rather not, (3) rather or (4) completely that they have (i) friendly classmates, (ii) caring classmates, (iii) many friends in class and (iv) no arguments with classmates. For each of the single questions, we define dummy variables that take the value one if the child rather agrees or completely agrees with the statement, and zero if the child does not or rather not agree with the statement. Accordingly, for the composite index, which we use as our main outcome for social integration, we define a dummy variable that takes the value one if the answers to the single question are mostly positive (i.e. if the mean of the answers is greater than 2.5) and zero if the answers to the single question are mostly negative (i.e. if the mean of the answers is 2.5 or less). In general, most answers indicate a high social integration. On average, non-native speakers agreed slightly less with the questions asked (see Table A1 in the Online Appendix).

Figure 2 shows the correlation between the three outcome variables and the linguistic diversity in the classroom, as measured by the Greenberg index. It reveals that linguistic diversity is negatively correlated with test scores with correlations of $-0.19$ to $-0.30$, but less so with the index of social integration where correlations range at around 0.1. Classes without any non-native speakers or with only one non-German language group have, by definition, a diversity of zero.

In the empirical analysis, we control for several individual and family characteristics to isolate the association between linguistic diversity and student outcomes. As individual characteristics, we include gender and a linear and quadratic measure of age, measured with monthly precision, to account for non-linear effects. To capture the pure age effect, we also include a dummy variable for students who repeated at least one of the four grades in primary school.

For non-native speakers, we further take the linguistic distance between their mother tongue and German into account. In addition, we include an indicator variable for whether these children are first-generation immigrants. As our samples are defined based on children’s mother tongue and not on their ancestry, native speakers may be born abroad or have foreign ancestry as well. We capture potential differences between these children and children without foreign ancestry by a dummy variable that equals one for first- or second-generation immigrants. For non-native speakers, we further add indicators for their region of origin, which capture the main so-called guest worker countries (Greece, Italy, Turkey), former Yugoslavia, Eastern European countries and the remaining countries.

Family characteristics cover the education level of the mother and the father in three categories (high, medium and low, based on the ISCED classification) as well as their employment status, distinguishing between white-collar workers, blue-collar workers and others. For mothers, we further control for whether they are full-time employed, part-time employed or not employed. In addition, we control for the number of books at home as a further proxy for students’ socio-economic background.

The school characteristics capture differences between cities of different size by a linear variable for the number of inhabitants at the school location and indicator variables for private vs. public and all-day vs. half-day schooling.

Our final sample includes 15,686 school children, containing 14,717 children whose mother tongue is German and 969 children whose mother tongue is a language other than German. These numbers show that our regression sample is considerably smaller than the raw sample of the survey. As in most educational surveys, non-response is therefore also an issue in our data. As can be seen from Online Appendix Table A2, which compares the the summary statistics of our sample to the unconditional statistics, non-response is not completely random. Our sample is slightly positively selected in a sense that the children included in our sample have, for example, better language and math skills and live in families with higher socio-economic background. However,
Figure 2. Correlation of linguistic diversity and student outcomes.
Source: Own calculations based on IQB 2011 data.
as the differences in socio-demographic characteristics between the two samples are rather small, we are confident that non-response does not alter our estimation results substantially.13

3. Empirical framework

To analyze the role of linguistic diversity in student performance, we estimate the following regression equation:

\[ y_{ic} = \alpha + \beta \text{migshare}_c + \gamma \text{diversity}_c + X_{ic}'\delta + S_c'\theta + \rho_s + \epsilon_{ic}, \]  

(4)

where \( y_{ic} \) denotes the test score or social integration index of student \( i \) in class \( c \). \( \text{migshare}_c \) is the share of students with a non-German mother tongue in class \( c \) and \( \text{diversity}_c \) is the linguistic diversity of non-native speakers in the class. \( X_{ic} \) are individual and family characteristics and \( S_c \) are school characteristics as described in Section 2. \( \rho_s \) denotes fixed effects for the 16 German federal states, which capture regional differences in population structures and student outcomes between the states. \( \epsilon_{ic} \) depicts the error term. All analyses are conducted separately for native and non-native speakers.14 Combined student, class, and school weights are used in all analyses and standard errors are clustered at the class level.

The main coefficient of interest is \( \hat{\gamma} \), the estimated effect of the linguistic diversity in the classroom on students’ test scores and social integration. For \( \hat{\gamma} \) to represent a causal estimate, we would have to assume that there are no unobserved characteristics that are both correlated with linguistic diversity and students’ outcomes. In the absence of panel data or a (quasi-)random allocation of students to schools and classes, this assumption is at risk of being violated. However, by conditioning on the share of non-native speakers in the classroom, we only explore variation in linguistic diversity across classes with similar levels of immigrant concentration. We argue that conditional on the share of non-native speakers in the class (as well as on an extensive set of background characteristics), non-random sorting of students into schools with a different linguistic composition of the immigrant group should not be a major concern. Given that the choice of primary school in Germany is mainly based on residency, families would have to take into account the linguistic diversity of their children’s peers in their location choice to let residential sorting threaten our identifying assumption. Put differently, although families might take into account the immigrant share in a district in their location choice, we argue that they do not explicitly consider the diversity level of the respective immigrant population in the school district.

Though we are not able to test the exogeneity assumption directly, we perform two indirect tests to check the validity of our identification assumption. Table 1 provides results from a balancing test, which analyzes whether observable characteristics that potentially influence student outcomes are correlated with linguistic diversity. Columns (1) and (3) contain the estimated coefficients from separate regressions for each control variable, i.e. they measure the unconditional correlation between linguistic diversity and each observable characteristic.15 Columns (2) and (4) show the results from similar regressions, but each of them conditioning on all other control variables, including the share of non-native speakers in the class. After including all other controls, the estimated coefficients should go to zero if variation in linguistic diversity is truly random. The results reveal that in the unconditional case, some of the background characteristics are correlated with linguistic diversity. However, they also show that including additional controls does a good job of removing the association between each specific control and linguistic diversity. Except for the share of non-native speakers in the class, which is – by definition – to some extent correlated with the diversity measure, hardly any of the other control variables are significantly correlated with linguistic diversity. In addition, the point estimates are very small in magnitude and the signs of the coefficients point in no clear direction. We thus conclude that, once we control for other observable characteristics, linguistic diversity is not systematically related to school and family background characteristics.
Table 1. Balancing test.

| Main explanatory variable | Native speakers | Non-native speakers |
|---------------------------|-----------------|-------------------|
| Share of non-native speakers in class | 131.09*** | 130.91*** |
|                           | (8.67)         | (9.24)            |
|                           | 62.81***       | 68.81***          |
|                           | (10.79)        | (10.62)           |
| Individual characteristics |                 |                   |
| Female                    | −1.04**        | 1.14              |
|                           | (0.47)         | (2.32)            |
| Age                       | −25.68*        | 43.09             |
|                           | (13.64)        | (57.35)           |
| Age²                      | 1.20*          | −1.97             |
|                           | (0.64)         | (2.69)            |
| Repeater                  | 1.02           | 5.03              |
|                           | (1.14)         | (3.93)            |
| 1st- or 2nd-generation migrant | 7.23***       | 0.59              |
|                           | (1.02)         |                   |
| 1st-generation migrant    | −1.24          | −1.28             |
|                           | (3.56)         | (2.98)            |
| Linguistic distance to German | −0.59**       | −0.81***          |
|                           | (0.28)         | (0.24)            |
| Region of origin (reference: main guest worker countries) | | |
| Former Yugoslavia         | −7.36*         | 2.69              |
|                           | (4.41)         | (3.57)            |
| Eastern Europe            | −0.64          | 3.47              |
|                           | (3.59)         | (3.34)            |
| Remaining (Europe and other continents) | 9.35**       | 8.92***           |
|                           | (4.32)         | (3.28)            |
| Family characteristics    |                 |                   |
| Number of books at home   | −0.00          | −0.04             |
|                           | (0.00)         | (0.02)            |
| Fathers’ education level (reference: low education) | | |
| Medium                    | −0.09          | 1.36              |
|                           | (1.04)         | (4.20)            |
| High                      | −0.30          | −1.04             |
|                           | (0.52)         | (2.35)            |
| Missing information       | −0.27          | 0.26              |
|                           | (0.96)         | (3.64)            |
| Mothers’ education level (reference: low education) | | |
| Medium                    | −0.08          | 10.62***          |
|                           | (0.90)         | (3.58)            |
| High                      | −1.16**        | 2.27              |
|                           | (0.53)         | (3.03)            |
| Missing information       | 0.75           | 7.34              |
|                           | (1.43)         | (4.47)            |
| Fathers’ employment status (reference: white-collar) | | |
| Blue-collar               | −0.85          | −1.75             |
|                           | (0.78)         | (3.89)            |
| Other                     | −1.48***       | 3.03              |
|                           | (0.56)         | (3.91)            |
| Missing information       | 0.44           | 4.59              |
|                           | (0.87)         | (4.17)            |
| Mothers’ employment status (reference: white-collar) | | |
| Blue-collar               | 1.63***        | 0.38              |
|                           | (0.80)         | (3.09)            |
| Other                     | −0.07          | −1.55             |
|                           | (0.56)         | (3.08)            |
| Missing information       | 0.70           | −3.02             |
|                           | (0.94)         | (3.49)            |
| Mothers’ labour force status (reference: full-time employed) | | |
| Part-time                 | 1.67***        | −2.96             |
|                           | (0.68)         | (3.22)            |
|                           | 1.07           | (3.14)            |

(Continued)
Nevertheless, to assess if the lack of complete balancedness might potentially skew results in a certain direction, we further perform an omnibus test. The test exploits the fact that, in order to bias our results, covariates would need to be systematically correlated with both linguistic diversity and students’ outcomes. In the first stage (results not shown), we use the full set of control variables, except for linguistic diversity, to predict our outcome variables, i.e. students’ language and math scores and their social integration. In the second stage, we then regress the predicted outcome variables on linguistic diversity. As Panel A of Table 2 shows, linguistic diversity is significantly negatively correlated with the three predicted outcome variables. However, this correlation accrues from the

Table 1. Continued.

| No controls | Controls | No controls | Controls |
|-------------|----------|-------------|----------|
| Native speakers | Coeff./SE | Coeff./SE | Coeff./SE | Coeff./SE |
| Non-native speakers | Coeff./SE | Coeff./SE | Coeff./SE | Coeff./SE |
| No employment | 1.64** | -0.45 | -4.16 | -1.60 |
| Missing information | 2.26 | 0.48 | -6.16 | -1.94 |

School and class characteristics

| Number of inhabitants at school location (in 1,000) | 0.03*** | -0.00 | 0.02** | -0.00 |
|---------------------------------------------------|---------|-------|---------|-------|
| Private school | -4.95** | -1.17 | -21.49*** | -29.69** |
| All-day school | 2.86* | 0.09 | -1.64 | -1.38 |
| State fixed effects | No | Yes | No | Yes |
| Classes | 1187 | 1187 | 460 | 460 |
| Observations | 14,717 | 14,717 | 969 | 969 |

Source: Own calculations based on IQB 2011 data.
Notes: OLS regression results with robust standard errors (clustered at class level) in parentheses. Columns (1) and (3) show the results from separate regressions of linguistic diversity on each explanatory variable (and a constant). Columns (2) and (4) show the results from a regression of linguistic diversity on all explanatory variables (and a constant). Asterisks indicate p-values according to: ***p < 0.01, **p < 0.05, *p < 0.1.

Nevertheless, to assess if the lack of complete balancedness might potentially skew results in a certain direction, we further perform an omnibus test. The test exploits the fact that, in order to bias our results, covariates would need to be systematically correlated with both linguistic diversity and students’ outcomes. In the first stage (results not shown), we use the full set of control variables, except for linguistic diversity, to predict our outcome variables, i.e. students’ language and math scores and their social integration. In the second stage, we then regress the predicted outcome variables on linguistic diversity. As Panel A of Table 2 shows, linguistic diversity is significantly negatively correlated with the three predicted outcome variables. However, this correlation accrues from the

Table 2. Omnibus test.

| Language test | Math test | Social integration |
|-----------|----------|------------------|
| Coeff./SE | Coeff./SE | Coeff./SE | Coeff./SE | Coeff./SE |

Panel A: No controls

| Native speakers | Coeff./SE | Coeff./SE | Coeff./SE | Coeff./SE |
|----------------|----------|----------|----------|----------|
| Non-native speakers | Coeff./SE | Coeff./SE | Coeff./SE | Coeff./SE |
| Linguistic diversity | -36.19*** | -34.55*** | -3.63*** | -12.08* | -33.32*** | -5.99*** |
| (4.87) | (5.23) | (0.62) | (6.86) | (9.76) | (2.23) |
| Classes | 1187 | 1187 | 1187 | 460 | 460 | 460 |
| Observations | 14,717 | 14,717 | 14,717 | 969 | 969 | 969 |

Panel B: Conditional on migrant share

| Share of non-native speakers in class | -169.76*** | -145.62*** | -16.77*** | -49.50*** | -74.31*** | -9.03** |
|--------------------------------------|------------|------------|------------|------------|------------|--------|
| (10.83) | (12.74) | (1.54) | (11.31) | (15.58) | (3.52) |
| Linguistic diversity | 7.36 | 2.82 | 0.51 | 4.60 | -8.28 | -2.95 |
| (4.82) | (6.19) | (0.67) | (7.52) | (9.87) | (2.17) |
| Classes | 1187 | 1187 | 1187 | 460 | 460 | 460 |
| Observations | 14,717 | 14,717 | 14,717 | 969 | 969 | 969 |

Source: Own calculations based on IQB 2011 data.
Notes: OLS regression results with robust standard errors (clustered at class level) in parentheses. The panels show second-stage results of regressions of predicted outcomes on linguistic diversity with no controls (Panel A) and the share of non-native speakers in the class as control variable (Panel B). For the sample of native speakers, a dummy variable for whether there are no non-native speakers in the class is further included in Panel B. In the first stage, outcomes are predicted based on all explanatory variables except for linguistic diversity. For ease of presentation, all coefficient estimates for social integration (columns (3) and (6)) are multiplied by 100. Asterisks indicate p-values according to: ***p < 0.01, **p < 0.05, *p < 0.1.
fact that in our setting, the share of non-native speakers is by definition positively correlated with linguistic diversity (and negatively correlated with the outcome variables). Once the share of non-native speakers is conditioned on (Panel B of Table 2), linguistic diversity is not significantly correlated with the predicted outcome variables. We are thus confident that observed (and unobserved) correlates of the linguistic composition of school classes are unlikely to confound our regression results.

4. Results

4.1. Baseline results

Tables 3 and 4 show the baseline results of estimating Equation (4) for the samples of native and non-native speakers. In each table, Columns (1) and (2) report the coefficient estimates for the share of non-native speakers, both unconditional and conditional on all control variables described in Section 2. In the following two columns, we extend the specification by including the linguistic diversity measure, again showing unconditional (Column (3)) as well as conditional (Column (4)) estimates. For ease of interpretation, all coefficient estimates for social integration (Panel C) are multiplied by 100.

For native speakers, the share of non-native speakers in the class is significantly negatively associated with both students’ test scores and their social integration. Conditional on all control variables (Column (2) of Table 3), a 10 percentage points increase in the share of non-native speakers in the class (which corresponds to an increase by approximately 1 standard deviation) lowers the language (math) test score of native speakers by around 8.7 (7.8) points, which corresponds to reductions in test scores of about 0.11 (0.09) standard deviations. The effects on language and math are therefore comparable in magnitude. A corresponding increase in the share of non-native speakers reduces the probability that native speakers are well integrated in the class by 1.1 percentage points or 0.04 standard deviations.

Table 3. Linguistic diversity and student outcomes – native speakers.

| Panel A: Language test score | Coeff./SE | Coeff./SE | Coeff./SE | Coeff./SE |
|-----------------------------|-----------|-----------|-----------|-----------|
| Share of non-native speakers in class | $-108.39^{***}$ | $-87.34^{***}$ | $-112.48^{***}$ | $-106.07^{***}$ |
| (15.89) | (15.21) | (24.67) | (23.38) |
| Linguistic diversity | $3.12$ | $-6.17$ | $(11.63)$ | $(10.61)$ |
| Control variables | No | Yes | No | Yes |
| Adjusted $R^2$ | 0.02 | 0.24 | 0.02 | 0.25 |

Panel B: Math test score

| Share of non-native speakers in class | $-97.58^{***}$ | $-77.66^{***}$ | $-95.59^{***}$ | $-93.63^{***}$ |
| (15.66) | (15.47) | (22.49) | (22.39) |
| Linguistic diversity | $-1.52$ | $-6.02$ | $(11.11)$ | $(10.32)$ |
| Control variables | No | Yes | No | Yes |
| Adjusted $R^2$ | 0.01 | 0.21 | 0.01 | 0.21 |

Panel C: Social integration

| Share of non-native speakers in class | $-10.57^{**}$ | $-11.29^{**}$ | $-17.34^{***}$ | $-18.22^{**}$ |
| (4.85) | (5.10) | (6.64) | (7.58) |
| Linguistic diversity | $-5.17$ | $4.63$ | $(3.28)$ | $(3.24)$ |
| Control variables | No | Yes | No | Yes |
| Adjusted $R^2$ | 0.00 | 0.03 | 0.00 | 0.03 |
| Classes | 1187 | 1187 | 1187 | 1187 |
| Observations | 14,717 | 14,717 | 14,717 | 14,717 |

Source: Own calculations based on IQB 2011 data.

Notes: – OLS regression results with robust standard errors (clustered at class level) in parentheses. Full estimation results of the model shown in column (4) are included in Table A4 in the Online Appendix. For ease of presentation, all coefficient estimates for social integration (Panel C) are multiplied by 100. Asterisks indicate $p$-values according to: $^{***}p < 0.01$, $^{**}p < 0.05$, $^*p < 0.1$. 
Given the share of non-native speakers in the class, the extent of linguistic diversity has no effect on the test scores or the social integration of native German speakers. Conditional on all other control variables (Column (4)), the coefficient estimates of linguistic diversity on students’ test scores are negative, but close to zero. For social integration, the respective coefficient estimate is positive, but again small and not statistically significant.

For non-native speakers, there is also a negative link between the share of non-native speakers in the class and students’ test scores (Column (1) of Table 4). The estimated coefficients are smaller than for the native sample, however, and not statistically significant for social integration. A 10 percentage points (1 standard deviation) increase in the share of non-native speakers is associated with a 4-point (0.1 standard deviations) reduction in the language test score and an 8-point (0.15 standard deviations) reduction in the math test score. In terms of standard deviations, the effect on language is therefore of comparable size to that for native speakers, while the effect on math is larger for non-native speakers.16

As for native speakers, the extent of linguistic diversity has no significant effect on the test scores of non-native speakers. For the language test score, the point estimate is negative and larger than for native speakers. The inclusion of the diversity index further reduces the coefficient estimate for the share of non-native speakers by about half. Therefore, both the share of non-native speakers and the degree of linguistic diversity among them have a non-negligible negative, but insignificant impact on the language proficiency of non-native speakers.

With respect to social integration, our results reveal that linguistic diversity significantly worsens the social integration of non-native speakers. An increase in linguistic diversity by 10 percentage points (1 standard deviation) lowers the probability that non-native speakers are well integrated in the class by 1.8 percentage points (0.13 standard deviations). Hence, our results reveal that the extent of linguistic diversity among the non-native speakers in a class is more important than the size of this group in determining the social integration of non-native speakers.
Table 5 further disentangles the diversity effect on non-native speakers by distinguishing between different components of the social integration index (columns (2) through (5)). The results reveal that a 10 percentage points increase in linguistic diversity increases the probability to have arguments with classmates by 1.1 percentage points. In addition, it decreases the probability to have many friends in the class by 1.9 percentage points. These results suggest that higher diversity might hamper communication among non-native speakers and could lead some students to feel isolated within the class. Whether classmates are friendly or caring in general, on the other hand, is less affected by the extent of linguistic diversity among classmates.

Our findings suggest that the degree of linguistic diversity neither helps nor hinders student performance as measured by test scores, but might challenge the social integration of non-native speakers in the class. The result that linguistic diversity is unrelated to students’ test scores is in line with Frattini and Meschi (2019), who, focusing on vocational training students, find no effect of ethnic diversity on natives’ test scores in math and literacy. It is, however, in contrast to the findings of Dronkers and van der Velden (2013), who find that ethnic diversity hampers the language skills of migrants (and of natives in highly stratified school systems). Maestri (2017), however, finds a positive effect of ethnic diversity on students’ test scores, in particular for language performance. Our small or null findings, and the different results in general, might be explained by two opposing effects working against each other: On the one hand, having many children with a similar mother tongue in the class may improve students’ self-consciousness through identity-building and foster communication and interaction among students with the same mother tongue. On the other hand, the formation of a large group of children with a different mother tongue than the native language may also slow the learning of this language and negatively impact learning by dividing the class and impeding the children’s sense of togetherness.

Table 5 in the Online Appendix shows the full estimation results of our preferred specification shown in Column (4) of Tables 3 and 4. Most individual and family characteristics are strong determinants of children’s school success, but explanatory power is higher for native speakers than for non-native speakers. The direction of effects is in line with theoretical expectations and in accordance with previous research.

4.2. Robustness checks

We perform several robustness checks in order to test whether the use of alternative measures of linguistic diversity, the inclusion of additional control variables, or changes in the sample affect our results. The respective regression results are summarized in Table 6.
Table 6. Robustness checks.

| Panel A: Linguistic polarization | Native speakers | Non-native speakers |
|----------------------------------|-----------------|---------------------|
| Share of non-native speakers in class | Language test score | Math test score | Social integration | Language test score | Math test score | Social integration |
|                                  | Coeff./SE       | Coeff./SE          | Coeff./SE         | Coeff./SE          | Coeff./SE          | Coeff./SE         |
| Linguistic polarization          | $-104.96^{***}$ | $-97.27^{***}$    | $-15.99^{**}$     | $-31.22$           | $-79.87^{***}$     | $-3.08$           |
|                                 | (21.05)         | (21.00)            | (7.21)            | (19.93)            | (21.53)            | (11.33)           |
| Linguistic polarization          | $-4.91$         | $-2.24$            | 2.03              | $-17.82^*$         | 7.74               | $-10.38^{**}$     |
|                                 | (5.95)          | (5.31)             | (1.80)            | (10.58)            | (9.77)             | (4.40)            |
| Adjusted $R^2$                   | 0.25            | 0.21               | 0.03              | 0.24               | 0.31               | 0.04              |
| Classes                          | 1187            | 1187               | 1187              | 460                | 460                | 460               |
| Observations                     | 14,717          | 14,717             | 14,717            | 969                | 969                | 969               |

| Panel B: Herfindahl-Hirschman index | Native speakers | Non-native speakers |
|-------------------------------------|-----------------|---------------------|
| Share of non-native speakers in class | Language test score | Math test score | Social integration | Language test score | Math test score | Social integration |
|                                    | Coeff./SE       | Coeff./SE          | Coeff./SE         | Coeff./SE          | Coeff./SE          | Coeff./SE         |
| Herfindahl-Hirschman index         | $-109.35^{***}$ | $-96.33^{***}$    | $-19.02^{**}$     | $-29.85$           | $-83.00^{***}$     | 2.68              |
|                                  | (23.12)         | (22.34)            | (7.53)            | (22.68)            | (22.30)            | (11.11)           |
| Herfindahl-Hirschman index         | 3.33            | 3.61               | $-4.81^*$         | 15.38              | $-10.10$           | 15.65**           |
|                                  | (9.45)          | (9.22)             | (2.89)            | (14.90)            | (13.69)            | (6.21)            |
| Adjusted $R^2$                    | 0.25            | 0.21               | 0.03              | 0.24               | 0.31               | 0.04              |
| Classes                           | 1187            | 1187               | 1187              | 460                | 460                | 460               |
| Observations                      | 14,717          | 14,717             | 14,717            | 969                | 969                | 969               |

| Panel C: Share of own language group | Native speakers | Non-native speakers |
|--------------------------------------|-----------------|---------------------|
| Share of non-native speakers in class | Language test score | Math test score | Social integration | Language test score | Math test score | Social integration |
|                                    | Coeff./SE       | Coeff./SE          | Coeff./SE         | Coeff./SE          | Coeff./SE          | Coeff./SE         |
| Share of own language group in class | $-32.82$         | $-63.23^{**}$     | $-18.40$          | $-23.40$           | $-34.09$           | 26.30              |
|                                    | (22.29)         | (24.84)            | (13.64)           | (43.21)            | (31.72)            | (16.71)           |
| Adjusted $R^2$                     | 0.23            | 0.31               | 0.04              | 0.23               | 0.31               | 0.04              |
| Classes                            | 1187            | 1187               | 1187              | 460                | 460                | 460               |
| Observations                       | 14,717          | 14,717             | 14,717            | 969                | 969                | 969               |

| Panel D: Cognitive skills          | Native speakers | Non-native speakers |
|------------------------------------|-----------------|---------------------|
| Share of non-native speakers in class | Language test score | Math test score | Social integration | Language test score | Math test score | Social integration |
|                                    | Coeff./SE       | Coeff./SE          | Coeff./SE         | Coeff./SE          | Coeff./SE          | Coeff./SE         |
| Linguistic diversity               | $-74.27^{***}$ | $-48.31^{***}$    | $-16.71^{**}$    | $-9.38$            | $-61.29^{***}$     | 4.58              |
|                                    | (21.38)         | (18.00)            | (7.54)            | (20.86)            | (20.37)            | (11.38)           |
| Cognitive skills                   | $-9.39$         | $-10.95$           | 4.33              | $-25.22$           | 3.96               | $-18.44^{***}$    |
|                                    | (9.32)          | (8.19)             | (3.21)            | (15.71)            | (13.44)            | (6.92)            |
|                                   | 5.02**          | 6.51***            | 0.21***           | 4.65***            | 5.29***            | 0.12              |
|                                   | (0.13)          | (0.15)             | (0.06)            | (0.41)             | (0.42)             | (0.25)            |
| Adjusted $R^2$                     | 0.40            | 0.42               | 0.03              | 0.37               | 0.45               | 0.04              |
| Classes                            | 1183            | 1183               | 1183              | 458                | 458                | 458               |
| Observations                       | 14,629          | 14,629             | 14,629            | 962                | 962                | 962               |

(Continued)
### Table 6. Continued.

|                        | Native speakers | Non-native speakers |
|------------------------|-----------------|---------------------|
|                        | Language test score | Math test score | Social integration | Language test score | Math test score | Social integration |
|                        | Coeff./SE     | Coeff./SE      | Coeff./SE      | Coeff./SE     | Coeff./SE      | Coeff./SE      |
| **Panel E: Migrant share at school** |                 |                   |                 |                 |                   |                 |
| Share of non-native speakers in class | $-95.80^{***}$ | $-84.42^{***}$ | $-23.66^{***}$ | $-1.19$ | $-72.69^{**}$ | $-3.25$ |
| (24.51) | (23.02) | (7.66) | (27.15) | (28.14) | (15.13) |
| Linguistic diversity | $-5.23$ | $-5.18$ | $4.14$ | $-19.38$ | $9.58$ | $18.46^{***}$ |
| (10.36) | (10.30) | (8.45) | (15.88) | (14.69) | (9.68) |
| Migrant share at school | $-12.89$ | $-11.57$ | $6.83^{***}$ | $-34.53$ | $-12.15$ | $9.08$ |
| (9.38) | (7.17) | (3.61) | (21.00) | (18.43) | (9.10) |
| Adjusted $R^2$       | 0.25            | 0.21             | 0.03            | 0.24            | 0.31            | 0.05            |
| Classes               | 1187            | 1187             | 1187            | 460            | 460            | 460            |
| Observations          | 14,717          | 14,717           | 14,717          | 969            | 969            | 969            |
| **Panel F: Exclude classes with only Turkish non-native speakers** |                 |                   |                 |                 |                   |                 |
| Share of non-native speakers in class | $-105.68^{***}$ | $-100.92^{***}$ | $-11.46$ | $-8.55$ | $-84.42^{***}$ | $8.28$ |
| (23.84) | (23.60) | (8.45) | (22.44) | (23.86) | (11.88) |
| Linguistic diversity | $-7.79$ | $-2.18$ | $4.99$ | $-21.29$ | $10.84$ | $19.72^{**}$ |
| (11.83) | (11.74) | (3.61) | (17.16) | (17.98) | (8.96) |
| Adjusted $R^2$       | 0.25            | 0.21             | 0.03            | 0.29            | 0.36            | 0.04            |
| Classes               | 1081            | 1081             | 1081            | 389            | 389            | 389            |
| Observations          | 13,518          | 13,518           | 13,518          | 827            | 827            | 827            |

**Source:** Own calculations based on IQB 2011 data.

**Notes:** OLS regression results with robust standard errors (clustered at class level) in parentheses. Panel A shows results similar to column (4) of Tables 3 and 4 where linguistic diversity is replaced by a measure for linguistic polarization, the Esteban-Ray polarization index (see Equation (5)). Panel B shows results similar to column (4) of Tables 3 and 4 where linguistic diversity is replaced by the Herfindahl-Hirschman index (see Equation (3)). Panel C shows results similar to column (4) of Tables 3 and 4 where linguistic diversity is replaced by the share of the own language group in the class. Panel D shows results similar to column (4) of Tables 3 and 4 where a measure for students’ cognitive skills is added as a further control variable. Panel E shows results similar to column (4) of Tables 3 and 4 where the migrant share at the school is added as control variable. Panel F shows results similar to column (4) of Tables 3 and 4 where classes with only German and Turkish speakers are excluded from the regressions. For ease of presentation, all coefficient estimates for social integration (columns (3) and (6)) are multiplied by 100. Asterisks indicate $p$-values according to: $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$. 


We start by using a measure of linguistic polarization instead of linguistic diversity to explain students’ outcomes. The so-called Esteban-Ray index (Esteban and Ray 1994) is calculated as

$$ER = \sum_{j=1}^{N} \sum_{k=1}^{N} s_j^{1+\alpha} s_k \delta_{jk},$$

(5)

where again, $N$ is the number of language groups with group shares $s_j$ and $\delta_{jk}$ is the linguistic distance between each pair of groups $j$ and $k$. As compared to the Greenberg index, the sensitivity factor $\alpha$ lets the index peak for two groups of the same size and maximum linguistic distance. In contrast to the Greenberg index, which monotonously increases in the number of different groups, the polarization index thus captures non-linear effects of a clustering of students along their native languages. The respective regression results using the polarization index are shown in Panel A of Table 6. Replacing linguistic diversity by linguistic polarization leaves the results for native speakers largely unaffected. For non-native speakers, the estimated negative effect of the linguistic composition in the classroom on language test scores is now more precisely estimated and turns significant at a 10-percent level. A 10 percentage points (1 standard deviation) increase in linguistic polarization reduces the language test scores of non-native speakers by 1.8 points (0.08 standard deviations). The respective effect of the share of non-native speakers remains insignificant. This result supports our previous finding for non-native speakers’ social integration, that the linguistic composition of the class is more important in explaining the outcome than the actual size of the group of non-German speakers. For math and social integration, the results remain largely unchanged.

In Panel B, we apply the Herfindahl-Hirschman index (Equation (3)) as an alternative measure for linguistic diversity, which has been used by previous studies that take linguistic or ethnic diversity in the class into account (Dronkers and van der Velden 2013; Maestri 2017; Frattini and Meschi 2019). As described in Section 2, the Herfindahl-Hirschman index is calculated as the sum of all squared language shares in a class. It thus differs from the Greenberg index in that the linguistic distance between each two groups is not considered. Moreover, the scale is reversed, i.e. the Herfindahl-Hirschman index decreases rather than increases in the number of different groups. The results show that for both native and non-native speakers, the results are largely robust to using this alternative measure of linguistic diversity. The main difference is that higher diversity now shows a positive impact on the social integration of native speakers. However, the respective point estimate is small and only significant at the 10-percent level. Overall, these results thus reveal that the inclusion of linguistic distance does not drive our estimation results.

Next, we test whether instead of the degree of linguistic diversity in general, the (relative) size of the own language group matters for schooling outcomes of non-native speakers. The results in Panel C, however, show that the share of students with the same mother tongue has no explanatory power for the test scores or the social integration of non-native speakers.

In the next step, we go back to our original model including linguistic diversity and add two further control variables. Panel D shows the results of including a measure of students’ cognitive skills in the model, which is obtained from a test containing deductive reasoning problems that all students had to solve as part of the survey. We do not include the results of this cognitive skill test in our main regressions, as they are closely correlated to test scores (here as well, the questions at hand have to be read and understood). The results reveal that cognitive skills are a strong predictor of students’ test scores, and that adding them to the model to some extent reduces the adverse impact of the share of non-native speakers in the class. The coefficient estimates for linguistic diversity, however, are hardly affected by controlling for students’ cognitive skills.

Panel E shows the results when adding the migrant share at the school level as an additional control variable. Controlling for this more aggregate share might capture neighborhood segregation above and beyond the share of non-native speakers in the class. The inclusion of this control variable reduces the coefficients for the share of non-native speakers in the class, because the two variables
are closely correlated. The results for the role of linguistic diversity in children’s schooling outcomes, however, remain unaltered.

Finally, we address the issue that the majority of non-native speakers represent children of Turkish decent. Thus, there could be a concern that many classes might only consist of German- and Turkish-speaking students, and that such classes are driving our results. In Panel F, we therefore drop classes in which Turkish-speaking students are the only non-German speakers in the class. The results are robust to the exclusion of these classes. The only difference that appears is that the negative effect of the share of non-native speakers in the class on the social integration of native speakers is no longer significant. However, the respective point estimate remains constant, but is now less precisely estimated due to the smaller sample size.

5. Conclusion

Rising immigration flows in many Western countries have led to an increase in the number of immigrant children in schools and changed the ethnic and linguistic composition of student populations. In this paper, we analyze the effect of immigrant peers in the classroom on the educational achievement and social integration of native and non-native speakers. While previous literature has mainly focused on investigating the effects of immigrant concentration in the class or school on student outcomes, we explicitly take the composition of the immigrant group into account. In doing so, we construct a novel measure of the degree of linguistic diversity in the class, which is based on contributions from the macroeconomic and political science literature (e.g. Easterly and Levine 1997; Alesina et al. 2003; Montalvo and Reynal-Querol 2005) and incorporates both the size of the different groups of non-native speakers and the linguistic distance between them. Our analysis makes use of a comprehensive survey of 4th-grade students in German primary schools, which contains detailed information on students’ migration background, family and school characteristics, results of standardized tests in both German language and math, as well as information on students’ social integration in the class. Germany represents an interesting case to analyze the effects of classroom diversity: As the assignment of children to primary schools is solely based on their residence, the share of immigrant students at a school is a consequence of the locational choice of families and can thus hardly be influenced. School principals, however, can regulate the allocation of migrants within a grade of a given school, thereby influencing the degree of linguistic diversity among immigrant students in a class.

A major identification problem when establishing the effect of class composition on student outcomes is related to student selection into schools. If schools with a relatively high linguistic diversity attract native and non-native children whose educational skills are different from those in schools with a relatively low linguistic diversity, we might erroneously conclude that diversity has spillover effects in the class. We address this potential selectivity by exploring the variation in linguistic diversity across classes with similar levels of immigrant concentration. Within the group of classes with the same share of non-native speakers, there may still be a selectivity issue due to potential non-random allocation of students to classes or higher allocation of teaching resource to classes with a higher linguistic diversity. However, based on balancing and omnibus tests, we find no evidence to support these concerns.

Our results reveal a negative association between the share of non-native speakers in the class and students’ test scores and their social integration. Conditional on the concentration of non-native speakers in the class, the degree of linguistic diversity, however, has no impact on students’ language and math test scores. This suggests that an increase in the number of students from more culturally or linguistically distant countries has no additional negative impact on students’ educational outcomes. We find though that a higher linguistic diversity in the class hampers the social integration of non-native speakers. In particular, non-native students in classes with a high linguistic diversity are more likely to have arguments with their classmates and have less friends in class. Hence, while the outcomes of native students are unaffected by the degree of
linguistic diversity, the social integration of non-native speakers could be improved by reducing linguistic diversity and allocating more students with the same linguistic background to the same class.

Notes

1. In 2017, the former German Secretary of Education, Johanna Wanka, demanded for the first time to put a cap on the number of migrant students in German classes, as a high share of migrant children in class would hamper the learning outcomes of all students (Frankfurter Allgemeine 2017).
2. For a cross-country comparative perspective, see, e.g. Dustmann, Frattini, and Lanzara (2012) and Brunello and Rocco (2013).
3. An exception is the study by Stanat (2006), which does not analyze diversity effects.
4. According to the German Microcensus, 28% of school children had a migration background in 2011. This share has been rising up to 36% in 2017 (Federal Statistical Office 2018).
5. While parents cannot freely choose a school for their children, residential sorting may, of course, still be a threat to our analysis. We more extensively discuss and test for this issue in Section 3.
6. A detailed description of the dataset is provided by Stanat et al. (2012, 2014) and Richter et al. (2014).
7. It is of course possible that children with a non-German mother tongue speak German at home. However, this would not be an issue for our analysis, as we do not argue that the non-native speakers have no language skills in German. Instead, for our analysis, it is relevant that these children also speak a different language than German, and that this language is the first language they have learned (i.e. their mother tongue).
8. Such imputation was done for 11% of the non-native speakers in our sample. Excluding these observations from the sample yields qualitatively and quantitatively similar results. For 170 children (15% of all non-native speakers), information on mother tongue could not unambiguously assigned based on country of birth, so that these children had to be excluded from the sample.
9. This equals the share of children with Turkish origin among all children with migration background aged 5 to 14 years obtained from official statistics (2011 German Microcensus; see Federal Statistical Office, 2018), revealing that our sample is representative with respect to the coverage of children with Turkish background.
10. Our results are further robust to excluding classes without non-native speakers from the sample.
11. The share of non-native speakers is smaller than the share of children with migration background, as many children speak German as a mother tongue even though their parents are immigrants (of all children in our sample who have at least one parent who was not born in Germany, 65% speak German as a mother tongue).
12. The distribution of the three dependent variables for the sample of native and non-native speakers is shown in Figure A2 in the Online Appendix.
13. Column 2 of Online Appendix Table A2 further reveals that the reduction in sample size is not due to a single variable that includes many missing values. Rather, each variable has a certain fraction of missing values, and these add up as more variables are included in the regression. This can be seen from Table A3 in the Online Appendix, which shows the number of observations when different groups of variables are step-wisely added. For example, even if only conditioning on non-missing values of all outcome variables, the sample is reduced by about 5000 observations or 18.6 percentage points (pp.). In addition, a considerable fraction of non-response arises from the fact that not all parents answered the parental questionnaire. Conditioning on the valid response to at least one question in the parental questionnaire reduces the sample by another 3000 observations (or 11.3 pp.). The parental questionnaire is crucial for our analysis, as information on children’s mother tongue, which is required for defining the sample and calculating our measure of linguistic diversity, is provided by the parents. Once conditioning on non-missing information on children’s mother tongue, non-response is not a major issue, as adding all remaining control variables reduces the sample by only 2400 observations (or 9.0 pp.).
14. Estimating a partially instead of fully interacted model yields similar results (see Table A5 in the Online Appendix).
15. For categorical variables, we include the indicator variables for all categories in one regression.
16. In terms of effect sizes, our estimates are larger than those found in previous analyses of the effect of the class share of immigrants on student performance (e.g. Ohinata and van Ours 2013; Diette and Oyelere 2014). However, as we are not able to fully address the problem of a selection of disadvantaged students into classes with many non-native speakers, our estimates represent upper bounds of the true effects of the share of non-native speakers on students’ outcomes.
17. Following previous literature (Montalvo and Reynal-Querol 2005), we set $\alpha$ to one and normalize the index to take on values between zero and one.
18. Note that due to the reverse scale of the Herfindahl-Hirschman index as compared to the Greenberg index, the signs of the coefficients are reversed.
19. This reduces the sample by 9% of classes for native speakers and by 15% of classes for non-native speakers.

Acknowledgements
The authors are grateful to Thomas Bauer, two anonymous reviewers, the editor Colin Green, and to participants of the 7th International Workshop on Applied Economics of Education, the 2017 annual meeting of the Leibniz Education Research Network (LERN), and the 2018 LERN Forum on Educational Policy for helpful comments and suggestions.

Disclosure statement
No potential conflict of interest was reported by the author(s).

Funding
This work was financially supported by the Mercator Foundation (project number 1534900).

References
Alesina, A., A. Devleeschauwer, W. Easterly, S. Kurlat, and R. Wacziarg. 2003. “Fractionalization.” Journal of Economic Growth 8 (2): 155–194. doi:10.1023/A:1024471506938. 
Arnold, D. H., and G. L. Doctoroff. 2003. “The Early Education of Socioeconomically Disadvantaged Children.” Annual Review of Psychology 54: 517–545. doi:10.1146/annurev.psych.54.111301.145442. 
Auerbach, A., and P. Oreopoulos. 1999. “Analyzing the Fiscal Impact of US Immigration.” American Economic Review 89 (2): 176–180. doi:10.1257/aer.89.2.176. 
Bakker, D., A. Müller, V. Velupillai, S. Wichmann, C. H. Brown, P. Brown, D. Egorov, R. Mailhammer, A. Grant, and E. W. Holman. 2009. “Adding Typology to Lexicostatistics: A Combined Approach to Language Classification.” Linguistic Typology 13 (1): 169–181. doi:10.1515/LITY.2009.009. 
Ballatore, R. M., M. Fort, and A. Ichino. 2018. “The Tower of Babel in the Classroom: Immigrants and Natives in Italian Schools.” Journal of Labor Economics 36 (4): 885–921. doi:10.1086/697524. 
BAMF. 2015. “Wanderungsmonitoring: Erwerbsmigration nach Deutschland.” Bundesamt für Migration und Flüchtlinge (BAMF), Jahresbericht. 
Borjas, G. J. 2003. “The Labor Demand Curve is Downward Sloping: Reexamining the Impact of Immigration on the Labor Market.” The Quarterly Journal of Economics 118 (4): 1335–1374. doi:10.1162/003355303322552810. 
Brunello, G., and L. Rocco. 2013. “The Effect of Immigration on the School Performance of Natives: Cross Country Evidence Using PISA Test Scores.” Economics of Education Review 32: 234–246. doi:10.1016/j.econedurev.2012.10.006. 
Card, D. 1990. “The Impact of the Mariel Boatlift on the Miami Labor Market.” Industrial and Labor Relations Review 43 (2): 245–257. doi:10.1177/001979399004300205. 
Card, D. 2001. “Immigrant Inflows, Native Outflows, and the Local Labor Market Impacts of Higher Immigration.” Journal of Labor Economics 19 (1): 22–64. doi:10.1086/209979. 
Casey, T., and C. Dustmann. 2008. “Intergenerational Transmission of Language Capital and Economic Outcomes.” Journal of Human Resources 43 (3): 660–687. doi:10.1353/jhr.2008.0002. 
Cho, R. M. 2012. “Are there Peer Effects Associated with Having English Language Learner (ELL) Classmates? Evidence From the Early Childhood Longitudinal Study Kindergarten Cohort (ECLS-K).” Economics of Education Review 31 (5): 629–643. doi:10.1016/j.econedurev.2012.04.006. 
Cornelissen, T., C. Dustmann, A. Raute, and U. Schönberg. 2018. “Who Benefits From Universal Child Care? Estimating Marginal Returns to Early Child Care Attendance.” Journal of Political Economy 126 (6): 2356–2409. doi:10.1086/699979. 
Desmet, K., I. Ortuño-Ortín, and S. Weber. 2005. “Peripheral Diversity and Redistribution.” CEPR Discussion Paper No. 5112. 
Desmet, K., I. Ortuño-Ortín, and S. Weber. 2009. “Linguistic Diversity and Redistribution.” Journal of the European Economic Association 7 (6): 1291–1318. doi:10.1162/JEEA.2009.7.6.1291. 
Diette, T. M., and R. U. Oyelere. 2014. “Gender and Race Heterogeneity: The Impact of Students with Limited English on Native Students’ Performance.” American Economic Review 104 (5): 412–417. doi:10.1257/aer.104.5.412. 
Diette, T. M., and R. U. Oyelere. 2017. “Do Limited English Students Jeopardize the Education of Other Students? Lessons From the North Carolina Public School System.” Education Economics 25 (5): 446–461. doi:10.1080/09645292.2017.1311300.
Dronkers, J., and R. van der Velden. 2013. “Positive but also Negative Effects of Ethnic Diversity in Schools on Educational Performance? An Empirical Test Using PISA Data.” In Integration and Inequality in Educational Institutions, edited by M. Windzio, 71–98. Cham: Springer.

Dustmann, C., and T. Frattini. 2014. “The Fiscal Effects of Immigration to the UK.” The Economic Journal 124 (S80): F593–F643. doi:10.1093/ecoj/12181.

Dustmann, C., T. Frattini, and G. Lanzara. 2012. “Educational Achievement of Second-Generation Immigrants: An International Comparison.” Economic Policy 27 (69): 143–185. doi:10.10111/j/1468-0327.2011.00275.x.

Dustmann, C., T. Frattini, and I. P. Preston. 2013. “The Effect of Immigration Along the Distribution of Wages.” The Review of Economic Studies 80 (1): 145–173. doi:10.1093/restud/rdq019.

Dustmann, C., U. Schönberg, and J. Stuhler. 2017. “Labor Supply Shocks, Native Wages, and the Adjustment of Local Employment.” The Quarterly Journal of Economics 132 (1): 435–483. doi:10.1093/qje/qjw032.

Easterly, W., and R. Levine. 1997. “Africa’s Growth Tragedy: Policies and Ethnic Divisions.” The Quarterly Journal of Economics 112 (4): 1203–1250. doi:10.1162/0033550555466.

Esteban, J., and D. Ray. 1994. “On the Measurement of Polarization.” Econometrica 62 (4): 819–851. doi:10.2307/2951734.

Federal Statistical Office. 2018. Bevölkerung mit Migrationshintergrund — Ergebnisse des Mikrozensus: Fachserie 1, Reihe 2.2.

Foged, M., and G. Peri. 2016. “Immigrants’ Effect on Native Workers: New Analysis on Longitudinal Data.” American Economic Journal: Applied Economics 8 (2): 1–34. doi:10.1257/app.20150114.

Frankfurter Allgemeine. 2017. “Union will Migranten-Grenze in Schulklassen.” Accessed 22 February 2019. https://www.faz.net/aktuell/politik/inland/integration-verbessemern-union-will-migranten-grenze-in-schulklassen-14993204.html.

Frattini, T., and E. Meschi. 2019. “The Effect of Immigrant Peers in Vocational Schools.” European Economic Review 113: 1–22. doi:10.1016/j.euroecorev.2018.12.005.

Giannelli, G. C., and C. Rapallini. 2016. “Immigrant Student Performance in Math: Does it Matter Where You Come From?” Economics of Education Review 52: 291–304. doi:10.1016/j.econedurev.2016.03.006.

Gould, E. D., V. Lavy, and M. D. Pascerman. 2009. “Does Immigration Affect the Long-Term Educational Outcomes of Natives? Quasi-Experimental Evidence.” The Economic Journal 119 (540): 1243–1269. doi:10.1111/j.1468-0297.2009.02271.x.

Greenberg, J. H. 1956. “The Measurement of Linguistic Diversity.” Language 32 (1): 109–115. doi:10.2307/410659.

Heckman, J. J. 2006. “Skill Formation and the Economics of Investing in Disadvantaged Children.” Science 312 (5782): 1900–1902. doi:10.1126/science.1128898.

Ispihording, I. E., and S. Otten. 2013. “The Costs of Babylon—Linguistic Distance in Applied Economics.” Review of International Economics 21 (2): 354–369. doi:10.1111/roie.12041.

Ispihording, I. E., and S. Otten. 2014. “Linguistic Barriers in the Destination Language Acquisition of Immigrants.” Journal of Economic Behavior & Organization 105: 30–50. doi:10.1016/j.jebo.2014.03.027.

Jensen, P., and A. W. Rasmussen. 2011. “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. doi:10.1016/j.econedurev.2011.08.002.

Kyriakides, L., and B. P. M. Creemers. 2008. “Using a Multidimensional Approach to Measure the Impact of Classroom-Level Factors Upon Student Achievement: A Study Testing the Validity of the Dynamic Model.” School Effectiveness and School Improvement 19 (2): 183–205. doi:10.1080/09243450802047873.

Lüdemann, E., and G. Schwerdt. 2013. “Migration Background and Educational Tracking.” Journal of Population Economics 26 (2): 455–481. doi:10.1007/s00148-012-0414-z.

Maestri, V. 2017. “Can Ethnic Diversity Have a Positive Effect on School Achievement?” Education Economics 25 (3): 290–303. doi:10.1080/09645292.2016.1238879.

Montalvo, J. G., and M. Reynal-Querol. 2005. “Ethnic Diversity and Economic Development.” Journal of Development Economics 76 (2): 293–323. doi:10.1016/j.jdeveco.2004.01.002.

Ohinata, A., and J. C. van Ours. 2013. “How Immigrant Children Affect the Academic Achievement of Native Dutch Children.” The Economic Journal 123 (S70): F308–F331. doi:10.1111/econo.12052.

Preston, I. 2014. “The Effect of Immigration on Public Finances.” The Economic Journal 124 (S80): F569–F592. doi:10.1111/econo.12180.

Richter, D., K. Böhme, J. Bastian-Wurzel, H. A. Pant, and P. Stanat. 2014. “IQB-Ländervergleich 2011.” Skalenhandbuch zur Dokumentation der Erhebungsinstrumente. Schriftenreihe des Institutes zur Qualitätsentwicklung im Bildungswesen No. 6.

Ruhose, J., and G. Schwerdt. 2016. “Does Early Educational Tracking Increase Migrant-native Achievement Gaps? Differences-in-Differences Evidence Across Countries.” Economics of Education Review 52: 134–154. doi:10.1016/j.econedurev.2016.02.004.

Schneeweis, N. 2015. “Immigrant Concentration in Schools: Consequences for Native and Migrant Students.” Labour Economics 35: 63–76. doi:10.1016/j.labeco.2015.03.004.

Spiegel Online. 2017. “Integration by the Numbers: Germany’s Ongoing Project to Welcome Its Refugees.” Accessed 21 December 2018. http://www.spiegel.de/international/germany/integrating-refugees-in-germany-an-update-a-1147053.html.
Stanat, P. 2006. “Schulleistungen von Jugendlichen mit Migrationshintergrund: Die Rolle der Zusammensetzung der Schülerschaft.” In *Herkunftsbedingte Disparitäten im Bildungswesen: Differenzielle Bildungsprozesse und Probleme der Verteilungsgerechtigkeit*, edited by J. Baumert, P. Stanat, and R. Watermann, 189–219. VS Verlag für Sozialwissenschaften.

Stanat, P., H. A. Pant, K. Böhme, and D. Richter, eds. 2012. *Kompetenzen von Schülerinnen und Schülern am Ende der vierten Jahrgangsstufe in den Fächern Deutsch und Mathematik: Ergebnisse des IQB-Ländervergleichs 2011*. Münster: Waxmann.

Stanat, P., H. A. Pant, K. Böhme, D. Richter, S. Weirich, N. Haag, A. Roppelt, M. Engelbert, and H. Reimers. 2014. IQB Ländervergleich Primarstufe 2011: Version: 2: Dataset. doi:10.5159/IQB_LV_2011_v2.

Swadesh, M. 1952. "Lexico-statistic Dating of Prehistoric Ethnic Contacts. With Special Reference to North American Indians and Eskimos." *Proceedings of the American Philosophical Society* 96 (4): 452–463.

The Economist. 2017. "Hearts and Minds: How Germany is Integrating Its Refugees." *The Economist* 424, 22–23.

Tonello, M. 2016. "Peer Effects of Non-native Students on Natives’ Educational Outcomes: Mechanisms and Evidence." *Empirical Economics* 51 (1): 383–414. doi:10.1007/s00181-015-0995-y.