Erasmus University Rotterdam

Erasmus School of Economics

Bachelor Thesis General Economics

The Influence of Ethnic Sorting on Educational Perfomance

Author
Tilbe Atav

Student number
408863

Supervisor
Prof. dr. Dinand Webbink

Second assessor
Matthijs Oosterveen

July 12, 2018

Abstract

This thesis investigates the link between the ethnic composition of classrooms in Dutch elementary schools and school results of non-Western minority children. Measures of ethnic classroom composition are set up to establish the concentration of non-Western minority children in general and the concentration of specific non-Western minority groups in classrooms. Using an individual fixed-effects approach, it is found that non-Western minority children perform better in classrooms with higher concentrations of children that do not have a Dutch ethnic background. Whether children are exposed to classrooms with a high share of children from their own ethnic background or not appears to be irrelevant.

1 Introduction

The gap between Dutch immigrants and non-immigrants persists. Dutch citizens from non-Western lineages are found to be unemployed more often, show lower levels of labour force participation, fall in lower income brackets more often and exhibit higher crime rates (CBS, 2016a; CBS, 2016b; CBS, 2016c). Research shows that this disparity starts at a young age: Dutch children with a non-Western background perform worse in the early years of elementary school than their peers with a fully Dutch background. Although a difference in school results seems to prevail in subsequent years, it is found to diminish towards the end of elementary school (Ledoux, Roeleveld, Driesen, Cuppen, & Meijer, 2011; Driessen, 2013). This shows that certain factors, to which children are subjected during this time period, decrease the gap between minority children and children with a Dutch background. Given that education is an essential determinant of cultural, social and socio-economic integration and the future position in the labour market of minorities, it is important to assess what these factors might be (CBS, 2016d). A significant influence on how children perform in school can stem from their peers and peer characteristics. Young children spend most of their day time in school, where they are in constant interaction with these peers. The influence of peers on learning and classroom processes can be positive as well as negative depending on peer characteristics (Hanushek, Kain, Markman, & Rivkin, 2003). Specifically, the influence of the ethnic background of peers on the performance of minority children will be considered in this research. The following research question is formulated:

What is the effect of the ethnic composition of classrooms in Dutch elementary schools on the school results of non-Western minorities?

Thus, the potential influence of shares of students from different backgrounds on test scores achieved by non-Western minority children will be analyzed. Specifically, the influence of the share of peers with a Dutch background and the influence of the share of peers with the same ethnic background will be investigated. This research focusses on children with a Turkish, Moroccan, Surinamese or Antillean background, as these are the largest non-Western communities in the Netherlands (CBS, 2016e).

A clear answer to the question posed above can be helpful in identifying the externalities that might arise from peer interaction. These externalities can be useful in determining optimal organization of a classroom (Hoxby, 2000). If certain structures of classroom composition diminish the gap between students from different ethnic backgrounds, it might be welfare-enhancing to implement sorting policies that enforce such compositions.

Although research on the influence of peers on student achievement is fairly common, most existing literature adresses the influence of peer achievements instead of the ethnic background of peers. Likewise, studies on educational inequality and on the school performance of minority students in the Netherlands have been conducted. However, investigation on how these concepts

relate to ethnic classroom composition remains scarce. This research is distinguished from existing literature by combining these directions to explore which relationships might hold for children in the Netherlands from specific ethnic backgrounds. The results show that being in a classroom with a higher share of non-Western minorities in general has a positive effect on the school results of non-Western minority children. However, whether a child from a specific ethnic background is exposed to a classroom composition that contains a high share of children that share this ethnic background or not seems to be unimportant in most cases.

In the following sections, steps will be taken towards answering the posed research question. First, in section 2, an overview of relevant literature is presented and expectations are formed. Subsequently, the methodology and data to test these expectations will be provided in section 3 and 4 respectively. In section 5, the results of the tests will be presented and described, followed by a robustness check in section 6. Finally, in section 7, the research is concluded.

2 Relevant literature

Relevant literature can be divided into two subsets: literature on peer effects and literature on ethnic sorting. In this section, examples of such research will be discussed.

Vast literature exists on the peer effects connected to classroom composition. These studies recognize and emphasize the direct and indirect effects that students have on their classmates. One of such studies is conducted by Hanushek, Kain, Markman and Rivkin (2003). These researchers perform an empirical analysis on peer effects, with which they tackle conceptual and data problems with regards to identifying the effect of peer group influences, such as omitted variable bias and the reciprocal nature of peer group interactions. Hanushek et al. do this by exploiting characteristics of panel data. They make use of controls for observed factors and allow for fixed individual, school and school-by-grade effects. Additionally, they use lagged measures of peer achievement to control for simultaneous equation bias. They find a positive influence of the achievement level of peers on learning outcomes: student results are higher if peers are high achievers (Hanushek, Kain, Markman, & Rivkin, 2003).

Other studies find positive relationships between peer achievement and learning outcomes as well (Hoxby, 2000; Zimmerman, 2003; Cárvo-Armengol, Patacchini, & Zenou, 2009; Jones, 2016). Studies previously performed in the Netherlands on this subject, however, find that this effect is very small (Roeleveld, Karssen, & Ledoux, 2014). Moreover, just like the study by Hanushek et al., these existing studies focus mainly on sorting with regards to student ability and not so much on what the influences of sorting with regards to ethnic background are. Some research *does* explore the dissimilarity of such effects for different racial groups and groups from various socio-economic backgrounds. These factors also seem to hold a significant relationship with student achievement (Winkler, 1975; Rumberger & Palardy, 2005). Additionally, research shows that a the relationship might vary among races and racial compositions of classrooms

(Hoxby, 2000; Winkler, 1975). Still, an analysis of ethnicity and specifically classroom composition in terms of ethnic background in this field is scarce.

This also applies to the field of study concerning the influence of sorting on inequality. A big concern within this line of research is that sorting might increase inequality. Most related studies focus on sorting in terms of educational level or income level, amongst other things. These studies assess whether sorting holds a relationship with educational inequality and income inequality. Although differences in magnitude exist, these studies often find a positive relationship (Kremer, 1997; Fernández & Rogerson, 2001; Fernández, Guner, & Knowles, 2005). However, the potential effects of ethnic sorting on educational inequality seem under-explored. The study that comes closest to the purpose of this research is one that is conducted by Friesen and Krauth (2007). Friesen and Krauth assess the relationship between, amongst others, ethnic sorting and educational inequality. They find that ethnic sorting is associated with higher levels of variance in educational results. The focus of the study performed by Friesen and Krauth differs from this research in several aspects, the main differences being in level of aggregation and in terms of country.

Based on the positive relationships found in existing literature between (ethnic) sorting and inequality, it is expected that ethnic sorting in elementary school classrooms is also positively related to inequality in school results of minority children. In oher words, as school results of non-Western minority children appear to be lower than the results of children with a Dutch background, these school results are expected to be even lower in combination with high levels of ethnic sorting in the classrooms. Additionally, previous research has shown that there can be differences in how children from different races are affected differently by peers. Therefore, it is expected that this difference will also prevail in terms of ethnicity.

3 Methodology

In this section, a description is provided of how the relationship between ethnic sorting in classrooms and school results of non-Western minorities will be evaluated to test whether the aforementioned expectations will hold. First, two measures of ethnic sorting are set up. After, the statistical methods that will be used to test the relationship will be discussed. A significance level of up to five percent will be used in assessing the statistical significance of all estimated coefficients.

3.1 Ethnic sorting

Two kinds of ethnic sorting are distinguished: ethnic sorting of minorities in general and ethnic sorting of specific groups of minorities. Both will be considered as there might be a difference between how the learning outcomes of a child from a certain ethnic background is affected by

being surrounded by children from a minority background in general and by children from the same ethnic background. As a measure of sorting of minorities in general, the percentage of children with a Dutch ethnic background is computed for each unique classroom c according to equation (1) presented below.

$$Percentage\,Dutch_c = \left(\frac{Number\,of\,children\,with\,Dutch\,parents_c}{Total\,number\,of\,children_c} \right) \cdot 100\% \tag{1}$$

As stated above, an additional measure will be set up to account for the effect of the classroom composition in terms of children with the same ethnic background. This measure is computed according to equation (2).

$$(Percentage\,not\,E_i)_c = \left(\frac{(Number\,of\,children\,not\,E_i)_c}{Total\,number\,of\,children_c} \right) \cdot 100\% \tag{2}$$

Similar to the previous measure, this measure is unique to each classroom c. Furthermore, this measure is unique to children from a certain ethnicity E, as their own ethnicity E_i is taken into account. The number of children that do not have a certain ethnicity E_i comprises of all children with parents that were not born in the country that coincides with ethnicity E_i.

3.2 Base model

To estimate the relationship between the ethnic composition of a classroom and school results, a linear regression where all observations are pooled is performed first. The following model is estimated:

$$Test\,score_i = \beta_0 + \beta_1 \cdot Percentage\,Dutch_{c_i} + \beta_2 \cdot S_{s_i} + \beta_j \cdot I_{ij} + \varepsilon_i \tag{3}$$

where the test score of child i is the dependent variable, the percentage of children with a Dutch ethnic background in the classroom c_i of child i is the independent variable and I_{ij} is a vector of control variables concerning child i. Furthermore, S_{s_i} is an indicator of the social-ethnic composition of the school attended by child i. A higher S_{s_i} coincides with a larger number of socio-ethnically disadvantaged children in a particular school. As mentioned in the introduction, this research focuses on the effects on the minority groups Turkish, Moroccan, Surinamese and Antillean specifically. Hence, the test scores of a child are regressed on the percentage of Dutch students in a classroom *only* for the children that have a Turkish, Moroccan, Surinamese or Antillean ethnic background. The estimate of coefficient β_1 ($\hat{\beta}_1$) can be interpreted as the change in the value of the test scores of these children with a one percentage point change in the percentage of Dutch children in a classroom.

Next, the second measure of ethnic sorting will be used. Likewise, a pooled linear regression

model is estimated:

$$Test\,score_{i,E} = \beta_0 + \beta_1 \cdot (Percentage\,not\,E_i)_c + \beta_2 \cdot S_{s_i} + \beta_j \cdot I_{ij} + \varepsilon_i \tag{4}$$

where again the test score of child i is the dependent variable. However, now, the percentage of children that do not have ethnic background E_i in the classroom c_i of child i is included as the independent variable. Another difference is that this model is estimated for all Turkish, Moroccan, Surinamese and Antillean children separately. The included control variables are the same. The estimate of coefficient β_1 ($\hat{\beta}_1$) can be interpreted as the change in the value of the test scores with a one percentage point change in the percentage of children in the classroom that do not have the same ethnic background as child i.

Using these models to investigate the effect of classroom composition is subject to several endogeneity issues. One of such issues is that some variables that are correlated with both the dependent and independent variables of interest are unobserved, for example factors behind how children are sorted into classrooms. Such variables cause omitted variable bias, as they are not included in or allowed for in the models set up above. Additionally, the test scores of children could have an effect on the classroom composition. Children may be retained or accelerated a year based on how they perform, for instance. This threat of reverse causality can be a source of bias too. How these issues are addressed in this research is discussed in more detail below.

3.3 Individual fixed-effects

The above described models pool the observations over either all children or all children with a certain ethnicity. Although some control variables are included to control for observed differences between these children, it is still highly likely that the coefficients are biased due to unobserved factors that might be correlated with both the ethnic composition of a classroom and the test scores of the children. It is then important to deliberate on where the variation in the dependent variables could come from.

As children themselves do not choose which school they attend or which classroom they are assigned to, two possible sources of variation stem from either the parents or the sorting policies of the school. If schools do not randomly assign children to a classroom but have a certain policy based on, for example, test scores, these policies would be correlated with both the test scores of a child and the ethnic composition of the classroom it is assigned to. This would cause the coefficients estimated with the aforementioned models to be biased. Additionally, a source of bias could be characteristics of the parents of a child. The parents of a child choose where they will reside, which holds relationship with the class composition that their child is subjected to. Moreover, parents' intelligence and stance towards education could influence both the area of residence and the test scores of their child. Altogether, this could lead to biased coefficients.

Assuming that these parental characteristics and school sorting policies are fixed over the

time period in which the children are observed, they can be accounted for by allowing for a child specific intercept as shown in specification (5).

$$Test\,score_{it} = \beta_i + \beta_1 \cdot Percentage\,Dutch_{c_{it}} + \beta_2 \cdot S_{s_i t} + \beta_j \cdot I_{ijt} + \varepsilon_{it} \tag{5}$$

Here, the β_i represents the child specific intercept for child i, which accounts for all child specific time-invariant factors that could be correlated with both the test scores of a child and the ethnic composition of the classroom. $S_{s,t}$ is an indicator of the social-ethnic composition of the school attended by child i at time t, and $I_{ij,t}$ represents a vector of j time-variant control variables concerning child i at time t. The same can be done for the effect of the percentage of children that do not share the ethnicity of child i, as presented in specification (6).

$$Test\,score_{it} = \beta_i + \beta_1 \cdot (Percentage\,not\,E_{it})_c + \beta_2 \cdot S_{s_i t} + \beta_j \cdot I_{ijt} + \varepsilon_{it} \tag{6}$$

Similar to before, specification (5) is estimated for all distinguished categories of non-Western minorities at the same time, while specification (6) is estimated for Turkish, Moroccan, Surinamese or Antillean children separately.

The coefficient of interest, $\hat{\beta}_1$, can be attained by eliminating the fixed child-specific differences, β_i, from the equation by taking the difference between the observed variables for child i at time t and the average value of these variables over all time periods for the same child i. This allows for a $\hat{\beta}_i$ to be obtained that is unbiased by fixed influences that cause differences between the test scores and ethnic composition of classrooms that each child is subjected to. Consequently, specification (7) can be estimated for the measure percentage of Dutch students.

$$Test\,score_{it} - \overline{Test\,score_{i_{5,8}}} = \beta_1 \cdot (Percentage\,Dutch_{c_{it}} - \overline{Percentage\,Dutch_{c_{i_{5,8}}}})$$

$$+\beta_2 \cdot (S_{s_i t} - \overline{S_{s_{i_{5,8}}}}) + \beta_j \cdot (I_{ijt} - \overline{I_{i_{5,8}j}}) + (\varepsilon_{it} - \overline{\varepsilon_{i_{5,8}}}) \tag{7}$$

Similarly, equation (8) is set up for the other measure of ethnic classroom composition.

$$Test\,score_{it} - \overline{Test\,score_{i_{5,8}}} = \beta_1 \cdot ((Percentage\,not\,E_{it})_c - \overline{(Percentage\,not\,E_{i_{5,8}})_c})$$

$$+\beta_2 \cdot (S_{s_i t} - \overline{S_{s_{i_{5,8}}}}) + \beta_j \cdot (I_{ijt} - \overline{I_{i_{5,8}j}}) + (\varepsilon_{it} - \overline{\varepsilon_{i_{5,8}}}) \tag{8}$$

An important assumption is made by interpreting the estimated $\hat{\beta}_1$ as the causal effect of of a certain measure of ethnic classroom composition on school results of non-Western minority children. This assumption is that no time-varying unobservables are correlated with both the used measure of ethnic classroom composition and the test scores of the included children at the same time. If there are such factors, these will be captured in the error term as unobserved

variables cannot be included in the models that are set up. Therefore, the estimated coefficient would be correlated with the error term, causing the coefficient to be biased as it will also include (some of) the effect of the time-varying omitted variables. Hence, the measured effect would not be the true effect of ethnic classroom composition on school results of non-Western minorities.

One such time-varying factor that could be a source of variation in the classroom composition is the fact that some children underperform to the degree where they have to be retained a year. This means that a child that has to retake the fifth year will not count towards the ethnic composition of the classroom of his or her peers when those peers are in their eighth year. As discussed earlier, non-Western minority children tend to score lower than children with a Dutch background. Therefore, it is reasonable to assume that they are more likely to repeat a year. Additionally, as will be discussed in the section below, there will be three years in between each observation for each child. This means it is more likely that one of the classmates of the child in question had to retake a year in either one of the years, compared to if there would be only one year between each observation. The children that will be included in estimating the above specified model, however, will only be the children that did manage to pass all included grades immediately. If this were not the case, they could not be matched. This rules out the possibility that their test scores were lower because they were already underperforming in the first place.

Regardless, there is potential for bias. Earlier it was mentioned that existing literature hints towards a small positive relationship between student achievement and achievement of peers (Hoxby, 2000; Hanushek, Kain, Markman, & Rivkin, 2003; Zimmerman, 2003; Cárvo-Armengol, Patacchini, & Zenou, 2009; Jones, 2016). Children that have to retake a year are naturally the children with lower test scores. Hence, it could be that their exclusion from the classroom has a positive effect on the achievement of the children that remain in it. This effect is also included in the estimated $\hat{\beta}_1$, yet it is not due to the change in the ethnic composition, but due to the change in the composition of the classroom in terms of the achievement of the peers of a child. Thus, the exclusion of a child from a classroom could create a bias in the estimated coefficients. For the percentage Dutch measure, this would cause an upward bias if the excluded child had to retake a year and, thus, had bad school results. However, if a child is excluded for other reasons, for example because the parents decided to move or if he or she was accelerated a year, it could also be that he or she had good school results, then this would cause a downward bias. For the other measure of ethnic composition of a classroom, this would be the other way around. Altogether, the expected bias from achievement based peer effects depends on the achievement of the children that are added to or removed from a classroom and is ambiguous for both measures.

4 Data

The data that will be analyzed to answer the research question is attained from the COOL$^{5\text{-}18}$ cohort studies, which mainly includes data on Dutch elementary school children in the school years 2007/2008, 2010/2011 and 2013/2014. From the year 2007/2008 and 2010/2011 information about children that are in their fifth year of elementary school will be extracted. From these waves, 12609 and 13266 observations are extracted respectively. Information about children that are in their eighth year of elementary school is also obtained. This data is attained from the academic years 2010/2011 and 2013/2014. Here, 12538 and 10058 observations were available. Altogether, this initial dataset contains 48471 observations. Specifically, data concerning the student number, school number, group number, group name, gender, weight-factor, test scores, place of birth of parents and the average test score of a school are gathered.

Using these data, several steps were taken to be able to estimate the effects of interest. First, the school number, group number, group name and a variable that indicates in which school year a specific child was in his or her fifth year of elementary school were combined to generate a variable that identifies each unique classroom. This was done for each child and for the fifth and eighth year separately. Subsequently, a number of dummy variables were created that indicate the ethnicity of a child. The ethnicities that are distinguished are Dutch, Turkish, Moroccan, Surinamese, Antillean and Other. To determine the ethnicity of a child, the place of birth of the parents are used: the dummies take the value one if either one or both of the parents of a specific child were born in the country that coincides with a specific ethnicity. The dummies were then used to determine the count of children with a certain ethnic background in each unique classroom, which in turn are used to compute the measures of ethnic composition.

As can be deduced from the posed research question, these measures of ethnic composition are two the main variables of interest within the context of this research. The other main variables concern the school results of each child in each year. The COOL$^{5\text{-}18}$ dataset contains test scores for, amongst others, reading and mathematics tests. These test scores lie within the same scale for each test, group and year, which allows for comparison. The mathematics and reading scores of a child will be used as measures of school results. All of the models set up in the previous section will be estimated for math scores and reading scores separately.

Additional to these main variables, a set of control variables will be used to account for observed differences between students. These include gender, wave, grade, test version, school score of the school attended and the OAB-weight factor. The wave variable indicates whether a child started their fifth year in 2007 or in 2011. In other words, it shows in which wave of the COOL$^{5\text{-}18}$ studies the child was observed for the first time. The grade variable indicates whether a child is in their fifth year or in their eighth year of elementary. The test version accounts for the fact that there are two versions of both the mathematics and the reading test, which are not comparable. Including it will make sure that only test scores of children that have made the same version are compared. The OAB-weight factor assigns a higher weight to students that are subject to disadvantages that could have a negative influence on their school

achievements, such as having lower educated parents. The school score variable indicates to what extent these disadvantaged children are present at a certain school.

As mentioned, the initially gathered data concerns 48471 observations. From these observations 43944 remain after all observations for children that were part of a mixed group[1] are dropped. After the measures of ethnic classroom composition are computed, all children that were not classified as Turkish, Moroccan, Surinamese or Antillean are dropped from the dataset as well. As a result, 7897 of the previous 43944 observations remain. Though initially classified as Turkish, Moroccan, Surinamese or Antillean to determine the ethnic classroom composition, some children only have one parent that was born in the related countries. As the parent with a Dutch background may cause the test scores of a child to be different from a fully non-Western background, these children are not included in the analysis. This concerns 1543 observations. Subsequently, observations that suffered missing values[2] or seemed to be subject to measurement errors[3] were dropped. After these steps, 5158 observations remain in the dataset. More precisely, the eventual dataset includes 2910 fifth year children and 2248 eighth year children. Of these children 2125 are Turkish, 2058 are Moroccan, 720 are Surinamese and 216 are Antillean. This pooled dataset will be used in estimating the base model.

To estimate the individual fixed effects model, on the other hand, the panel characteristics of the obtained data will be exploited. A substantial share of the children that are included have been observed in both their fifth year of elementary as well their eighth year. The data about these children that have been measured on two different points in time can be matched to be able to control for the fixed influences on their test scores. After matching, 906 unique children remain in the dataset, which are observed in both the fifth and the eighth year. Thus, in total, the dataset contains 1812 observations, which means that 3725 observations have been dropped from the pooled dataset, as not all children were observed in two consecutive time periods. Of the remaining 1812 observations 818 are Turkish, 712 are Moroccan, 214 are Surinamese and 68 are Antillean.

4.1 Descriptive statistics

A summary of the data is presented in table 1. A comparison between the values for the pooled and matched data sets shows that the differences between the two data sets is relatively small, which implies that the matched data is representative of the initial, larger set of observations. This implication is supported by the estimations of the binary logistic regression in Appendix table A.1. To attain these estimations, a dummy, which takes value 1 for all observations in

[1]Some schools make use of mixed groups, where children from different years of elementary are pooled together. For these groups, only information about the children that are in their fifth or eight year is included in the COOL$^{5\text{-}18}$ studies, which means it is not possible to determine the full class-composition. Hence, classrooms for which the group name implied that a group was mixed, were dropped from the dataset.

[2]Missing values occured mostly in the test scores, but could be found in the other used variables as well.

[3]For example, some children were classified as a different gender in two consecutive years.

the pooled dataset and value 0 for all observations in the matched dataset, is regressed on certain characteristics of the included children and the schools they attend. The estimated coefficients are mostly insignificant, showing that having these these characteristics do not increase (or decrease) the probability of being in the pooled dataset as opposed to the matched dataset. The constant, however is positive and significant, which is not surprising as the pooled dataset is larger than the matched dataset. The only remaining significant coefficient belongs to the dummy Turkish, which takes value 1 for children with a Turkish background and zero for children with a different background. The probability that can be computed with this coefficient is positive, which implies that being Turkish increases the probability of being in the pooled dataset[4]. This indicates that Turkish children are relatively overrepresented in the pooled dataset as compared to the matched dataset.

Table 1: Descriptive statistics for Turkish, Moroccan, Surinamese and Antillean children

	Pooled		Matched	
Variable	Mean	SD	Mean	SD
Math test	84.241	25.480	86.885	25.125
Reading test	30.295	19.511	32.181	19.343
Percentage not Turkish	79.293	17.726	77.570	18.637
Percentage not Moroccan	78.037	20.983	79.693	20.404
Percentage not Surinamese	91.375	17.743	93.006	15.300
Percentage not Antillean	98.018	4.460	97.735	4.418
Percentage Dutch	24.006	26.068	25.810	26.151
Weight-factor	0.620	0.522	0.630	0.520
Gender	0.505	0.500	0.520	0.500
Wave	0.529	0.499	0.573	0.495
Grade	1.458	0.498	1.500	0.500

The descriptives presented in table 1 can provide insight to the composition and distribution of the data sets. For example, the results for the math test appear to be less volatile compared to the results for the reading test. This implies that the differences in the reading test scores are higher between students and grades than the differences in math test results. Additionally, there are large differences in the average values and standard deviations of the measures of ethnic composition. The average percentage of children that are Dutch is relatively low, but is more volatile. The measures for the percentage that is not Surinamese or Antillean on the other had are much higher and are less volatile. The values for the measures for percentage of children that are not Turkish or Moroccan lie in between. Lastly, there is a fairly equal division

[4]Compared to being Antillean, as the dummy variable Antillean, which takes value 1 for children with an Antillean background and 0 for children with a different background, was ommitted due to perfect multicollinearity.

of the possible genders, waves and grades.

Furthermore, correlation-matrices can give an initial idea of how the test scores and the measures of ethnic composition relate to each other. These can be found in Appendix tables A.2 (pooled) and A.3 (matched). Again, the values for the two data sets are fairly similar. In both cases, the percentage not Turkish and the percentage not Antillean hold a weak positive relationship with both test scores, whereas the percentage not Moroccan and the percentage not Surinamese hold a weak negative relationship. Additionally, in both cases the test scores are strongly positively correlated with each other. This is means that children that do well in maths also tend to do well in reading. Finally, one difference between the two matrices is that for pooled data the percentage of Dutch children holds a weak negative relationship with both test results, whereas for the matched dataset the coefficient is positive for the reading test score. However, in both cases the correlation between the percentage of Dutch children and reading test scores is very close to zero.

Lastly, to be able to estimate the fixed-effects models, it is important that there are differences between the ethnic compositions in classrooms over time for a child. To assess how much classroom composition varies between the two points in time, a difference between these two values can be taken for each child. Figure A.1 in the Appendix shows these differences for the ethnic classroom composition measure percentage of Dutch children for each child in the matched data set. Although there are some children that do not experience any difference, a large part of the children do. Most of the values are fairly close to the line where the difference is zero or fluctuate around it up to a difference of twenty percent. Some outliers, however, can be found at a difference of forty percent. As mentioned in the previous section, the regular levels of variation can have three sources: school policy, parental characteristics, and entry or exit of children. The outliers could concern children that moved and joined a school that is also part of the COOL$^{5\text{-}18}$sample, or it could be that schools merged or split, causing higher variation in class composition. Similarly, Appendix figure A.2 shows the differences for the other measure of ethnic classroom composition for each ethnicity separately. It is not surprising that, again, there is variability. However, the measures are less volatile for children with a Surinamese or Antillean background. These ethnic groups are also subject to a much lower number of observations.

5 Results

Below, the results of both the pooled base model and the fixed effects model are presented and discussed. In this section, only the coefficients of the measures of ethnic composition of classrooms are included in the tables. More elaborate tables, which include coefficients for the control variables, can be found in the Appendix tables A.4-7.

5.1 Base model

First, the mathematics test scores were regressed on the measures of ethnic classroom composition. Table 2 shows the estimated coefficients. The first column contains the result for model (1). Columns two to five contain the results for model (2) for children with a Turkish, Moroccan, Surinamese and Antillean ethnic background respectively. As shown by the adjusted R-squared, all five specifications explain sixty to seventy percent of the variation in the math test scores.

The coefficients estimated in the first column show that there is a small, statistically significant, negative relationship between math test scores and the percentage of children with a Dutch ethnic background in a classroom. More specifically, when the percentage of children with a Dutch ethnic background in a classroom increases by one percentage point, the math test scores of a child with a non-Western ethnic background decreases by 0.034 points. Thus, children with a non-Western background tend to perform worse in math in classrooms with a larger share of children with a Dutch background.

Table 2: Pooled regression results for the relationship between ethnic composition measures and math test results

	Math test scores				
	All	Turkish	Moroccan	Surinamese	Antillean
Percentage Dutch	-0.034*				
	(0.014)				
Percentage not Turkish		0.086***			
		(0.024)			
Percentage not Moroccan			0.011		
			(0.018)		
Percentage not Surinamese				-0.093***	
				(0.015)	
Percentage not Antillean					0.154
					(0.157)
Observations	5158	2125	2058	720	216
Adjusted R-squared	0.635	0.615	0.641	0.716	0.616

Standard errors in parentheses; * p<0.05, ** p<0.01, *** p<0.001; each column contains the result of an OLS-regression where the math test scores of a certain group of children is regressed on a measure of ethnic composition of a classroom and several control variables; the first column shows the result for the regression of math test scores of all non-Western minority children on the percentage of Dutch children; the second to fifth columns show the results for the regression of the math test scores of Turkish, Moroccan, Surinamese and Antillean children are regressed on the percentage children in a classroom that are not Turkish, Moroccan, Surinamese or Antillean respectively; the full results can be found in Appendix table A.4.

A similar significant relationship can be found for Surinamese children in column four. When the percentage of children with an ethnic background that is not Surinamese increases with

one percentage point, the math test results of Surinamese children decreases by 0.093 points. This means that a child with a Surinamese background tends to perform worse in classrooms with less children that also have a Surinamese background. For Turkish children, on the other hand, this relationship seems to be significant and positive. As shown by column two, when the percentage of children with an ethnic background that is not Turkish increases by one percentage point, the math test results of Turkish children increases by 0.086 points: Turkish children tend to perform better in classrooms with fewer other Turkish children. For Moroccan and Antillean children, however, the relationship seems to be insignificantly different from zero. These children appear to be unaffected by changes in the percentage of children with a non-Moroccan or non-Antillean ethnic background in the classroom.

Subsequently, the reading test scores were regressed on the measures of ethnic classroom composition. Table 3 shows the estimated coefficients specifically for the reading tests. Again, the first column contains the result for model (1). Columns two to five contain the results for model (2) for children with a Turkish, Moroccan, Surinamese and Antillean ethnic background respectively. As shown by the adjusted R-squared, all five specifications explain approximately fifty percent of the variation in reading test scores.

Table 3: Pooled regression results for the relationship between ethnic composition measures and reading test results

	Reading test scores				
	All	Turkish	Moroccan	Surinamese	Antillean
Percentage Dutch	-0.048***				
	(0.012)				
Percentage not Turkish		0.078***			
		(0.020)			
Percentage not Moroccan			-0.017		
			(0.019)		
Percentage not Surinamese				-0.066***	
				(0.017)	
Percentage not Antillean					-0.020
					(0.118)
Observations	5158	2125	2058	720	216
Adjusted R-squared	0.471	0.462	0.491	0.508	0.471

Standard errors in parentheses; * p<0.05, ** p<0.01, *** p<0.001; each column contains the result of an OLS-regression where the reading test scores of a certain group of children is regressed on a measure of ethnic composition of a classroom and several control variables; the first column shows the result for the regression of reading test scores of all non-Western minority children on the percentage of Dutch children; the second to fifth columns show the results for the regression of the reading test scores of Turkish, Moroccan, Surinamese and Antillean children are regressed on the percentage children in a classroom that are not Turkish, Moroccan, Surinamese or Antillean respectively; the full results can be found in Appendix table A.5.

Here, the results are almost identical to the math test results in terms of sign and significance.

The effects do, however, seem to differ in magnitude. The relationship appears to be stronger in the first column, which regards the percentage of children with a Dutch background. This means that the reading test scores of children with a non-Western background tend to be affected more by changes in the percentage of Dutch children in a classroom compared to their math scores. For column two and four, which concern the second measure of ethnic classroom composition specifically for Turkish and Surinamese children respectively, the relationship is slightly weaker compared to the relationship that was found in the estimations for the math scores. This means that the reading test scores of these children are affected less by changes in the percentage of children with a different ethnic background than their mathematics test scores are.

5.2 Individual fixed-effects

Subsequently, the same relationships are tested again. Now, however, the fixed child-specific intercepts are eliminated from the equations. The resulting coefficients are presented in tables 4 and 5 for math and reading test results respectively. Each column contains the estimates for one of the measures of ethnic classroom composition. These are presented in the same order as in the tables above.

As mentioned, the results of the effect of ethnic composition of classrooms on math test scores can be found in table 4. The estimations for the first ethnic composition measure are shown in the first column. Compared to the estimates of the base model for mathematics in table 2, the coefficient for the first measure of ethnic composition of classrooms (column one) has decreased, whereas the standard error is relatively smaller. This shows that the unobserved fixed influences that are still present in the base model caused an upward biased coefficient. However, after this bias is eliminated, a similar relationship holds: children with a non-Western ethnic background tend to do worse in classrooms with higher shares of children with a Dutch background. Specifically, when the percentage of children with a Dutch background increases by one percentage point, the math test results of children with a Turkish, Moroccan, Surinamese or Antillean background decreases by 0.130 points.

The estimates of the second measure of ethnic classroom composition, presented in columns two to five in table 4, however, do show larger differences in comparison with the base model. Where children with a Turkish (column two) or Surinamese (column four) ethnic background were subject to a statistically significant coefficients before, they are not after allowing for fixed differences between each child from a certain ethnicity. In the fixed-effects estimations, the effect of having more or less children with the same ethnic background on math test scores seems to be insignificantly different from zero for Turkish, Moroccan and Surinamese children. Surprisingly, the coefficient for the effect of the percentage of children with a non-Antillean background *is* significant. More precisely, when the percentage of children with a non-Antillean background increases by one percentage point, the math test score of an Antillean child decreases by 1.713 points, which implies that Antillean children seem to perform better in terms of mathematics

in classrooms with fewer other Antillean children. Important to note is that the coefficient is much higher compared to all other statistically significant results. Additionally, the number of observations for Antillean children is fairly low

Table 4: Individual fixed-effects regression results for the relationship between ethnic composition measures and math test results

	Math test scores				
	All	Turkish	Moroccan	Surinamese	Antillean
Percentage Dutch	-0.130**				
	(0.039)				
Percentage not Turkish		-0.032			
		(0.087)			
Percentage not Moroccan			0.021		
			(0.110)		
Percentage not Surinamese				-0.215	
				(0.147)	
Percentage not Antillean					1.713***
					(0.273)
Observations	1812	818	712	214	68
Within R-squared	0.877	0.853	0.886	0.942	0.963
Between R-squared	0.026	0.013	0.007	0.074	0.074
Overall R-squared	0.563	0.379	0.562	0.685	0.635

Standard errors in parentheses; * p<0.05, ** p<0.01, *** p<0.001; each column contains the result of an OLS-regression where the math test scores of a certain group of children is regressed on a measure of ethnic composition of a classroom and several control variables; the first column shows the result for the regression of math test scores of all non-Western minority children on the percentage of Dutch children; the second to fifth columns show the results for the regression of the math test scores of Turkish, Moroccan, Surinamese and Antillean children are regressed on the percentage children in a classroom that are not Turkish, Moroccan, Surinamese or Antillean respectively; the full results can be found in Appendix table A.6.

Similar to the math test results, the estimates for the measure that concerns the percentage of children that have a different ethnic background than the child in question are insignificantly different from zero for Turkish, Moroccan and Surinamese children. These estimates are presented in columns two to four. In column five the result for the same measurement is shown for Antillean children. Here, the coefficients do differ significantly from the math test results: where the percentage of non-Antillean children held a large, positive and statistically significant relationship with the math test results, it is insignificantly different from zero for reading test results.

Another similarity with the math test results is the statistically significant, negative relationship between the reading test scores and the percentage of children with a Dutch background in a classroom. The standard error of this coefficient, however, is higher relative to the coefficient in question. This coefficient shows that the reading test scores of Turkish, Moroccan, Surinamese

and Antillean children will decrease by 0.095 points with a one percentage point increase in the percentage of children with a Dutch ethnic background in the classroom. Thus, children with a non-Western ethnic background tend to perform worse in classrooms with a higher share of children with a Dutch background. Again, the coefficient of this measure seems to have become lower after eliminating the fixed influences on the reading test scores of minority children. This means that the base model for reading test scores was subject to upward bias as well.

Table 5: Individual fixed-effects regression results for the relationship between ethnic composition measures and reading test results

	Reading test scores				
	All	Turkish	Moroccan	Surinamese	Antillean
Percentage Dutch	-0.095*				
	(0.047)				
Percentage not Turkish		-0.164			
		(0.085)			
Percentage not Moroccan			-0.050		
			(0.078)		
Percentage not Surinamese				0.143	
				(0.213)	
Percentage not Antillean					0.143
					(0.461)
Observations	1812	818	712	214	68
Within R-squared	0.794	0.799	0.800	0.771	0.836
Between R-squared	0.000	0.021	0.006	0.044	0.123
Overall R-squared	0.429	0.449	0.423	0.346	0.575

Standard errors in parentheses; * p<0.05, ** p<0.01, *** p<0.001; each column contains the result of an OLS-regression where the reading test scores of a certain group of children is regressed on a measure of ethnic composition of a classroom and several control variables; the first column shows the result for the regression of reading test scores of all non-Western minority children on the percentage of Dutch children; the second to fifth columns show the results for the regression of the reading test scores of Turkish, Moroccan, Surinamese and Antillean children are regressed on the percentage children in a classroom that are not Turkish, Moroccan, Surinamese or Antillean respectively; the full results can be found in Appendix table A.7.

Lastly, both table 4 and 5 include values for the within and between R-squared of each specification. Noteworthy is that for all specifications the within R-squared is fairly high with values between seventy and ninety percent. This means that, depending on the specification, seventy to ninety percent of the variation in school results between the fifth and eighth year of elementary school for a child with a non-Western ethnic background is explained by the variables included in the specifications. The between R-squared of each specification, however, seems to be substantially lower and in some cases even close to zero. This means that a very low share of the differences in the test result between different children is explained by these models. This means that, although the estimations may give a good image of the influences

on the math and reading test scores of minority children within their fifth and eighth year of elementary, differences in the math and reading test scores between children remains fairly unexplained. For the purposes of this research, this does not pose a problem, as the main focus is on the influence of a change in classroom composition on a particular child.

5.3 Other interesting results

Additional to the main results that are discussed above, there are a couple of other interesting findings. One noteworthy observation is that the coefficients of the grade variable are strongly significant as shown in Appendix tables A.4 to A.7. The estimated coefficients are quite similar for the base models and the fixed effect models. The largest source of variation in the test results is captured with this variable. The value of this variable shows how non-Western minority children develop their mathematics and reading skills between the fifth and eighth year. The estimates for the fixed-effects model show that non-Western minority children in general witness an increase of 41.300 points in math test scores and an increase of 26.480 points in reading test scores. Comparing these values to the values for the grade coefficients of specific ethnicities shows how ethnicities differ in how they develop. The results in Appendix table A.6 show that Antillean children experience a lower increase in their mathematics test scores compared to other non-Western minority children. The estimates in Appendix table A.7, on the other hand, show that Antillean children experience a higher increase in their reading test scores compared to other non-Western minority children, whereas Surinamese children show a lower increase.

Furthermore, the values for the school score variable are zero for Surinamese and Antillean children in the fixed-effects models (Appendix A.6 and A.7), indicating that the schools of these children fell into the same interval in terms of school scores when these children were in their fifth and in their eighth year. As the value of the variable is therefore time-invariant, a fixed-effects estimation method will not estimate a coefficient. For column one and two, however, the coefficients are estimated and even statistically significant (Appendix A.7). These results make sense: non-Western minority children in general and specifically Turkish children seem to perform worse in schools where the student population tends to consists of more children that are disadvantaged in socio-ethnic terms. For Moroccan children a coefficient is also estimated. However, this coefficient is statistically insignificant.

6 Robustness to outliers

To estimate the results discussed above, Ordinary Least Squares (OLS) regressions were performed. With this method the models are estimated such that the squared distance between observations are minimized given a set of variables. This means that the estimated model give a distorted image of the general trend in most observations if a small set of datapoint significantly deviate from this trend. To test the sensitivity of the above discussed results to

such outliers, these models can be estimated once more after eliminating these outliers from the dataset. Specifically outliers in the independent variables of interest will be considered. This will only be done for the individual fixed-effects models, as the base models have shown to be subject to bias.

The box-plots in Appendix figure A.3 show that the four dependent variables that result from the second measure of ethnic classroom composition, namely percentage not Turkish, percentage not Moroccan, percentage not Surinamese and percentage not Antillean, are subject to some clear outliers. For the first measure of ethnic classroom composition, namely percentage Dutch, however, the box-plot does not show any observations that are situated at a significant distance from the rest. To still be able to identify observations that are in the extremes of the distribution z-scores will be computed. Subsequently, benchmarks are chosen. Z-scores that fall above and below these benchmarks will be identified as possible outliers in the measures for ethnic composition of classrooms. Lastly, these outliers are eliminated from the dataset after which the models can be estimated.

The benchmark z-scores that are chosen in this research are 2.5 and -2.5. These are chosen as the percentage Dutch measure only has z-scores between 3 and -3. Removing observations with a z-score below -2.5 and above 2.5 will ensure that the observations that are left in the dataset are the ones within the interval of z-scores where 98.8 percent of all observations lie. Estimating the models set up above after eliminating the identified outliers gives the estimations presented in Appendix table A.8 for the math scores and Appendix table A.9 for the reading scores. Comparing these estimations to tables A.6 and A.7 of the Appendix, which hold the results of the previous individual fixed-effects estimations, shows how many observations were deleted.

The estimates show minor differences for all measures and specifications. The coefficients of the main variables of interest in the specifications regarding the second measure of ethnic composition experience negligible changes in magnitude and remain insignificant for both test scores. Similar observations can be made about the coefficient of the percentage Dutch variable in the estimates for the math test score. The coefficient of the percentage Dutch variable in the estimates for the reading test score, on the other hand, seems to become statistically insignificant as the coefficient decreases while the standard error remains the same. However, it is important to note that the decrease was fairly small and that the coefficient was nearly insignificant at a significance level of five percent to begin with. Additionally, looking at Appendix figure A.4, one can observe that actually no observations deviate significantly from the distribution of the rest of the observations. This raises the question whether these observations should in fact be removed from the dataset.

All in all, eliminating outliers from the dataset does not appear to significantly impact the coefficients that are estimated and hence also do not change the main findings of this research.

7 Conclusion

The previous sections have focussed on investigating the question: *What is the effect of the ethnic composition of classrooms in Dutch elementary schools on the school results of non-Western minorities?* Here, the findings of this research with regards to this question will be discussed. In section five, allowing for fixed differences between the children showed that the estimates of the base model were subject to bias. Therefore, only the estimates that result from the fixed effects model are used in answering the posed research question.

To estimate the relationship between the ethnic classroom composition and school results of non-Western minorities, two different measures for this composition were used. The first measure is the percentage of children in a classroom with a Dutch background and the second measure is the percentage of children that do not have a certain ethnic background. The latter was used to test whether sorting of more or less children with the same non-Western ethnicity in a certain classroom would have an effect on the how children from that specific ethnicity perform. For this measure, almost no significant effects were found. The other measure was included to test whether the sorting of more or less children from all included non-Western ethnic backgrounds in a classroom would affect the test scores of these children. Here, a small but significant negative relationship was found between classrooms with higher shares of children with a Dutch background and the school results of non-Western minority children. In other words, non-Western minority children seemed to perform better when sorted into classrooms with more other non-Western minority children. Moreover, for this measurement, it was found that mathematics test scores are affected to a greater extent than the reading test scores. Correcting for outliers did not seem to change these results.

All in all, the effect of the ethnic composition of a classroom differs per measure. Being in a classroom with a higher share of non-Western minorities in general seems to have a positive effect on the school results of non-Western minority children. However, whether a child from a specific ethnic background is exposed to a classroom composition that contains a high share of children that share this ethnic background or not appears to be unimportant in most cases.

Looking back at existing literature these findings seem to be unconventional. More specifically, the expectations derived from relevant literature in section 2 stated the exact opposite from what was eventually found. In particular, the results concerning the effect on skills related to language seem counterintuitive. The opposite result was expected, given that, classrooms with higher shares of children with a Dutch background children were expected to be in contact with the Dutch language more often.

These surprising relationships that were found could stem from several sources. First of all, it could be that a child performs better at school if he or she is surrounded by children that are similar to him or her. It might be that non-Western minority children share certain characteristics, which allow them to feel more comfortable and at home in classrooms with other non-Western minorities. Such characteristics could have an impact on how easily children befriend each other and how a child feels about going to school, whereas being surrounded by

children that are very different in terms of culture, behavior and school performance could make a child feel isolated and arouse negative feelings towards going to school and participating. Some support for this explanation based on peer relationships can be found in existing literature. For example, research has shown that peer interactions have a significant influence on motivation through prosocial behavior, school adjustment and emotional distress (Wentzel, 1998). Furthermore, group membership has been found to be a significant determinant of school results (Wentzel & Caldwell, 1997).

Additionally, the pace in the classrooms and teacher performance could play a role in why these results were found (Roeleveld, Karssen & Ledoux, 2014). As mentioned before, minority children usually start with a set back compared to children with a Dutch background. It could be that the pace in classrooms is dependent on how a majority of the class performs, which might be problematic for students that cannot keep up. Furthermore, the participation and results of children that perform below average may depend on the ability of a teacher to give special attention to these children. Thus, these factors could play a role in the difference between how children perform in classrooms with different composition.

Furthermore, a possible explanation for these results could be sought out in the characteristics of the children with a Dutch background and how they vary over classrooms with a different ethnic composition. For example, characteristics of children with a Dutch background in classrooms with higher percentages of these children may be more often factors that can be a negative influence on the school results of non-Western minority children. A characteristic that could be of influence, for instance, is to what extent a child is found to be disadvantaged. However, patterns in the data show that children with a Dutch background seem to be less disadvantaged on average in classrooms with a larger share of ethnically Dutch children, which seems to contradict this idea[5].

Another explanation behind these unconventional results could be that, notwithstanding the elimination of fixed-effects, the results could be biased by factors that are time-variant over the three years in which the children were observed. As discussed in section 3, this bias could stem from peer effects in terms of changes in peer achievement levels, due to for example the entry or exit of students from a particular classroom (Hoxby, 2000; Hanushek, Kain, Markman, & Rivkin, 2003; Zimmerman, 2003; Cárvo-Armengol, Patacchini, & Zenou, 2009; Jones, 2016). However, a bias due to peer achievement can cause the results to be the opposite of the expectations only if the net change in the average achievement level in a classroom of children leaving and entering is positively related to the share of non-Western minorities in a classroom. In other words, the non-Western children that enter the classroom have to increase the average achievement level or children that have to retake due to low school results need to be the children with a Dutch ethnic background. If the opposite is true, then the bias would be upward and hence the actual coefficient would be even lower, which would

[5]A figure that shows the distribution of the average weight factor of Dutch children in a classroom, which is an indicator of disadvantage in terms of parental education, over classrooms with different percentages of children with a Dutch background is presented in Appendix A.5.

still lead to an effect with the same unexpected sign. Given that it was stated earlier that children with a Dutch background tend to perform better, however, it seems unlikely that this criterium will hold. Still, this potential for bias is one of the limitations of this research.

Several other limitations should be addressed. First of all, as mentioned in section two, any peer effects can be reciprocal in nature: a specific child is affected by its peers but the peers are also affected by this specific child. Hanushek et al. (2003) attempt to solve this problem in their research by using lagged terms of their variable of interest. However, given the data at hand, this was not possible here. Furthermore, the measures of ethnic classroom composition are expressed in percentages. Therefore, changes in composition of different sizes of classrooms might be misrepresented. For example, if fifty percent of the minority children leave a classroom where there were only two minority children, it is weighted equally to if fifty percent of the minority children leave a classroom where initially all children had a non-Western ethnicity. However it could be that, in actually, these effects are different. Another limitation is that there are three years between the each observation for a child, which increases the risk for time-variant omitted variable bias. Especially given the fact that little observed time-variant factors were available. Furthermore, the data set contains little observations over time per child. Lastly, many observations needed to be dropped for the various reasons mentioned in section 4.

Future research could focus resolving these limitations and reassessing the research question. This could be done by making use of data that has more observations in the time or cross-sectional dimensions or by using different measures of classroom composition. Furthermore, this research focusses only on particular minority groups. How the investigated relationships hold for children with a Dutch background or children from other minority groups remains unknown. Thus, further research could assess the overall effect of ethnic classroom composition. Additionally, peer effects might be more prominent in later years of education where peer influences like peer pressure become more important. Therefore, an interesting area of further research is the effect of ethnic classroom composition in later years of education, for example in high school.

For now, the relationships that were found by this investigation imply that a higher concentration of non-Western minority children in a classroom would be beneficial for their school results. However, where this relationship comes from exactly remains unknown. Besides, given the magnitude of the found effects, the estimated models have explained little to no difference between the results of children with a Dutch ethnic background and children with a non-Western ethnic background. Given that ethnic composition of classrooms does not seem to be a substantial influence on how minority children perform in school, clearly other factors are at play that drive the wedge between children from different ethnic backgrounds. The quest to understanding why a gap between children with and without an immigration background exists, therefore, continues.

References

[1] Cárvo-Armengol, A., Patacchini, E., & Zenou, Y. (2009). Peer Effects and Social Networks in Education. The Review of Economic Studies, 76(4), 1239–1267.

[2] CBS. (2016a). Arbeidsparticipatie. Retrieved from CBS: https://www.cbs.nl/nl-nl/achtergrond/2016/47/arbeidsparticipatie

[3] CBS. (2016b). Criminaliteit. Retrieved from CBS: https://www.cbs.nl/nl-nl/achtergrond/2016/47/arbeidsparticipatie

[4] CBS. (2016c). Inkomen. Retrieved from CBS: https://www.cbs.nl/nl-nl/achtergrond/2016/47/inkomen

[5] CBS. (2016d). Onderwijs. Retrieved from CBS: https://www.cbs.nl/nl-nl/achtergrond/2016/47/onderwijs

[6] Driessen, G. (2013). De ontwikkeling van de onderwijspositie van de doelgroepen van het onderwijsachterstandenbeleid tussen 1995 en 2011. Tijdschrift voor Orthopedagogiek, 52, 67-81.

[7] Fernández, R., & Rogerson, R. (2001). Sorting and Long-Run Inequality. The Quarterly Journal of Economics, 116(4), 1305–1341.

[8] Fernández, R., Guner, N., & Knowles, J. (2005). Love and Money: A Theoretical and Empirical Analysis of Household Sorting and Inequality. The Quarterly Journal of Economics, 120(1), 273–344.

[9] Friesen, J., & Krauth, B. (2007). Sorting and inequality in Canadian schools. Journal of Public Economics, 91, 2185–2212.

[10] Hanushek, E. A., Kain, J. F., Markman, J. M., & Rivkin, S. G. (2003). Does peer ability affect student Achievement? Journal of Applied Econometrics, 18, 527-544.

[11] Hoxby, C. (2000). Peer effects in the classroom: Learning from gender and race variation. NBER Working Paper 7867, National Bureau of Economic Research.

[12] Jones, S. (2016). How does classroom composition affect learning outcomes in Ugandan primary schools? International Journal of Educational Development, 48, 66-78.

[13] Kremer, M. (1997). How Much does Sorting Increase Inequality? The Quarterly Journal of Economics, 112(1), 115–139.

[14] Ledoux, G., Roeleveld, J., Driessen, G., Cuppen, J., & Meijer, J. (2011). Doelgroepleerlingen in het basisonderwijs: historische ontwikkeling en actuele situatie. Amsterdam: Kohnstamm Instituut.

[15] Roeleveld, J., Karssen, M., & Ledoux, G. (2014). Samenstelling van de klas en cognitieve en sociaal-emotionele uitkomsten. Amsterdam: Kohnstamm Instituut.

[16] Rumberger, R. W., & Palardy, G. J. (2005). Does segregation still matter? The impact of student composition on academic achievement in high school. Teachers College Record, 107(9), 1999-2045.

[17] Wentzel, K. R. (1998). Social Relationships and Motivation in Middle School: The Role of Parents, Teachers, and Peers. Journal of Educational Psychology, 90(2), 202-209.

[18] Wentzel, K. R., & Caldwell, K. (1997). Friendships, Peer Acceptance, and Group Membership: Relations to Academic Achievement in Middle School. Child Development, 68(6), 1198-1209.

[19] Winkler, D. R. (1975). Educational achievement and school peer group composition. The Journal of Human Resources, 10(2), 189-204.

[20] Zimmerman, D. J. (2003). Peer Effects in Academic Outcomes: Evidence from a Natural Experiment. Review of Economics and Statistics, 85(1), 9-23.

Appendix

Table A.1: Logistic regression of dataset dummy on child and school characteristics

	Dataset
Weight factor	0.007
	(0.056)
Gender	-0.075
	(0.057)
Turkish	-0.294*
	(0.144)
Moroccan	-0.173
	(0.145)
Surinamese	-0.025
	(0.157)
Grade	-0.088
	(0.090)
Reading score	-0.003
	(0.002)
Math score	-0.003
	(0.002)
Wave	-0.046
	(0.056)
Constant	1.855***
	(0.189)

Standard errors in parentheses; * p<0.05, ** p<0.01, *** p<0.001; the dependent variable is a dummy that takes value one for all individuals in the pooled dataset and value zero for all individuals in the matched dataset.

Table A.2: Correlation-coefficients between test results and measures of ethnic composition (Pooled data)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
(1) Math	1.000						
(2) Reading	0.736*	1.000					
(3) Share not Turkish	0.018	0.086*	1.000				
(4) Share not Moroccan	-0.054*	-0.036*	0.029*	1.000			
(5) Share not Surinamese	-0.048*	-0.095*	-0.327*	-0.290*	1.000		
(6) Share not Antillean	0.090*	0.083*	-0.120*	-0.178*	0.128*	1.000	
(7) Share Dutch	-0.044*	-0.011	0.389*	0.526*	0.226*	0.041*	1.000

* p<0.05

Table A.3: Correlation-coefficients between test results and measures of ethnic composition (Matched data)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
(1) Math	1.000						
(2) Reading	0.754*	1.000					
(3) Share not Turkish	0.006	0.069*	1.000				
(4) Share not Moroccan	-0.012	-0.021	-0.097*	1.000			
(5) Share not Surinamese	-0.021	-0.049*	-0.298*	-0.279*	1.000		
(6) Share not Antillean	0.054*	0.099*	-0.079*	-0.224*	0.272*	1.000	
(7) Share Dutch	-0.016	0.004	0.389*	0.494*	0.219*	0.078*	1.000

* p<0.05

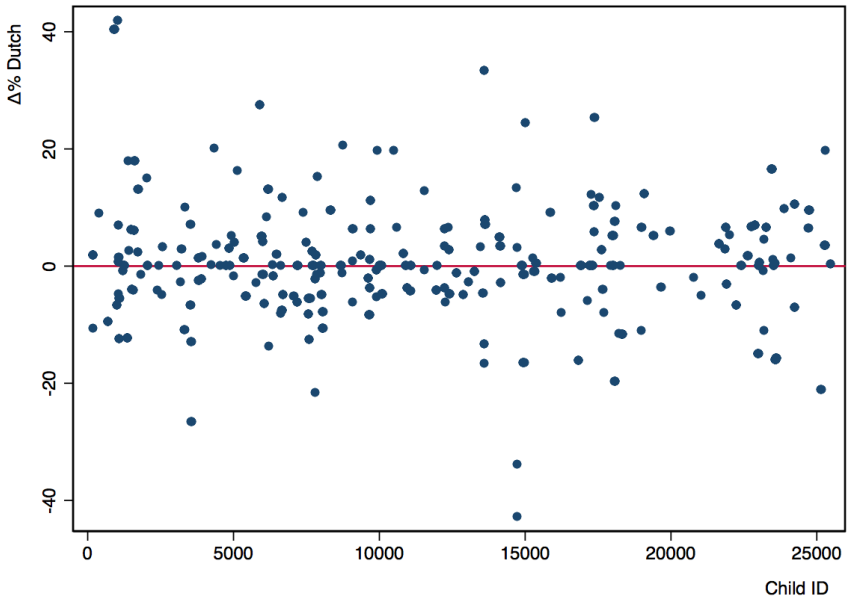

Figure A.1: Change in percentage of children with a Dutch ethnic background between measurement in fifth year and eighth year for each child

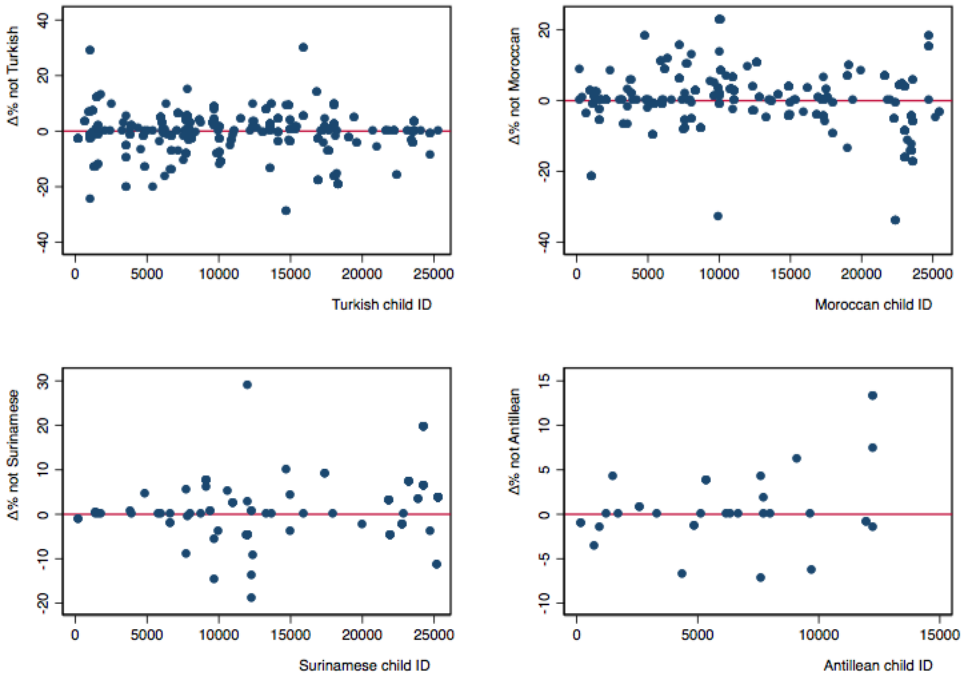

Figure A.2: Change in percentage of children with a different ethnic background between measurement in fifth year and eighth year for each child from a specific ethnic background

Table A.4: Pooled regression results for the relationship between ethnic composition measures and math test results

	Math test scores				
	All	Turkish	Moroccan	Surinamese	Antillean
Percentage Dutch	-0.034*				
	(0.014)				
Percentage not Turkish		0.086***			
		(0.024)			
Percentage not Moroccan			0.011		
			(0.018)		
Percentage not Surinamese				-0.093***	
				(0.015)	
Percentage not Antillean					0.154
					(0.157)
Version	-12.720***	-13.390***	-12.990***	-9.268***	-11.440***
	(0.506)	(0.748)	(0.815)	(1.390)	(2.599)
Grade	40.880***	40.400***	41.310***	40.340***	38.270***
	(0.464)	(0.758)	(0.724)	(1.025)	(2.544)
Gender	-5.219***	-5.884***	-5.368***	-2.942**	-7.135***
	(0.428)	(0.684)	(0.674)	(0.982)	(1.923)
Weight-factor	-0.968*	-1.077	-2.033**	-0.871	-1.606
	(0.436)	(0.721)	(0.684)	(1.299)	(2.525)
School score	-0.146	0.440**	0.341*	-0.749***	0.301
	(0.120)	(0.136)	(0.157)	(0.165)	(0.348)
Wave	0.669	2.115**	0.175	-0.173	-0.198
	(0.496)	(0.802)	(0.773)	(1.105)	(3.093)
Constant	39.990***	28.710***	35.550***	46.520***	20.770
	(1.539)	(2.776)	(2.551)	(2.830)	(15.750)
Observations	5158	2125	2058	720	216
Adjusted R-squared	0.635	0.615	0.641	0.716	0.616

Standard errors in parentheses; * p<0.05, ** p<0.01, *** p<0.001; each column contains the result of an OLS-regression where the math test scores of a certain group of children is regressed on a measure of ethnic composition of a classroom and several control variables; the first column shows the result for the regression of math test scores of all non-Western minority children on the percentage of Dutch children; the second to fifth columns show the results for the regression of the math test scores of Turkish, Moroccan, Surinamese and Antillean children are regressed on the percentage children in a classroom that are not Turkish, Moroccan, Surinamese or Antillean respectively.

Table A.5: Pooled regression results for the relationship between ethnic composition measures and reading test results

	Reading test scores				
	All	Turkish	Moroccan	Surinamese	Antillean
Percentage Dutch	-0.048***				
	(0.012)				
Percentage not Turkish		0.078***			
		(0.020)			
Percentage not Moroccan			-0.017		
			(0.019)		
Percentage not Surinamese				-0.066***	
				(0.017)	
Percentage not Antillean					-0.020
					(0.118)
Version	-3.525***	-5.032***	-2.295*	-4.076**	2.086
	(0.570)	(0.869)	(0.934)	(1.543)	(2.200)
Grade	25.340***	24.900***	26.430***	24.810***	24.350***
	(0.421)	(0.653)	(0.653)	(1.201)	(2.438)
Gender	3.222***	3.036***	3.463***	4.369***	-1.202
	(0.394)	(0.614)	(0.613)	(1.034)	(1.617)
Weight-factor	-3.230***	-2.390***	-4.407***	-0.570	-1.530
	(0.407)	(0.659)	(0.634)	(1.345)	(2.368)
School score	-0.515***	0.014	-0.102	-0.697***	-0.130
	(0.113)	(0.120)	(0.142)	(0.166)	(0.328)
Wave	0.604	1.407*	0.215	-1.538	0.609
	(0.425)	(0.660)	(0.655)	(1.165)	(2.277)
Constant	1.911	-10.380***	-1.950	9.556**	-3.268
	(1.539)	(2.447)	(2.837)	(3.271)	(12.730)
Observations	5158	2125	2058	720	216
Adjusted R-squared	0.471	0.462	0.491	0.508	0.471

Standard errors in parentheses; * p<0.05, ** p<0.01, *** p<0.001; each column contains the result of an OLS-regression where the reading test scores of a certain group of children is regressed on a measure of ethnic composition of a classroom and several control variables; the first column shows the result for the regression of reading test scores of all non-Western minority children on the percentage of Dutch children; the second to fifth columns show the results for the regression of the reading test scores of Turkish, Moroccan, Surinamese and Antillean children are regressed on the percentage children in a classroom that are not Turkish, Moroccan, Surinamese or Antillean respectively.

Table A.6: Individual fixed-effects regression results for the relationship between ethnic composition measures and math test results

	Math test scores				
	All	Turkish	Moroccan	Surinamese	Antillean
Percentage Dutch	-0.130**				
	(0.039)				
Percentage not Turkish		-0.032			
		(0.087)			
Percentage not Moroccan			0.021		
			(0.110)		
Percentage not Surinamese				-0.215	
				(0.147)	
Percentage not Antillean					1.713***
					(0.273)
Grade	41.300***	41.510***	41.330***	41.510***	35.370***
	(0.574)	(1.004)	(0.819)	(1.131)	(2.588)
Weight-factor	-0.1284	-1.925	-1.385	-2.229	-5.328
	(1.113)	(1.870)	(1.630)	(3.498)	(3.025)
Version	-13.890***	-15.280***	-13.970***	-12.300***	-14.830***
	(1.206)	(2.130)	(1.990)	(3.361)	(2.861)
School score	-3.828***	-6.511***	-2.323	0	0
	(0.940)	(0.949)	(1.241)	(.)	(.)
Constant	73.010***	96.680***	55.070***	48.910***	-119.100***
	(8.751)	(10.310)	(15.540)	(10.890)	(24.860)
Observations	1812	818	712	214	68
Within R-squared	0.877	0.853	0.886	0.942	0.963
Between R-squared	0.026	0.013	0.007	0.074	0.074
Overall R-squared	0.563	0.379	0.562	0.685	0.635

Standard errors in parentheses; * p<0.05, ** p<0.01, *** p<0.001; each column contains the result of an OLS-regression where the math test scores of a certain group of children is regressed on a measure of ethnic composition of a classroom and several control variables; the first column shows the result for the regression of math test scores of all non-Western minority children on the percentage of Dutch children; the second to fifth columns show the results for the regression of the math test scores of Turkish, Moroccan, Surinamese and Antillean children are regressed on the percentage children in a classroom that are not Turkish, Moroccan, Surinamese or Antillean respectively.

Table A.7: Individual fixed-effects regression results for the relationship between ethnic composition measures and reading test results

	Reading test scores				
	All	Turkish	Moroccan	Surinamese	Antillean
Percentage Dutch	-0.095*				
	(0.047)				
Percentage not Turkish		-0.164			
		(0.085)			
Percentage not Moroccan			-0.050		
			(0.078)		
Percentage not Surinamese				0.143	
				(0.213)	
Percentage not Antillean					0.143
					(0.461)
Grade	26.480***	26.990***	26.710***	23.920***	28.410***
	(0.480)	(0.733)	(0.734)	(1.717)	(2.590)
Weight-factor	-0.838	-0.480	-3.721*	-1.081	-5.766
	(1.030)	(1.553)	(1.650)	(3.738)	(4.846)
Version	-1.865*	-2.329*	-1.022	-1.762	-0.494
	(0.744)	(1.114)	(1.236)	(2.481)	(2.200)
School score	0.776	-0.316	2.159	0	0
	(1.482)	(1.356)	(2.346)	(.)	(.)
Constant	-9.810	-6.833	-19.850	-9.519	-27.240
	(13.150)	(12.980)	(22.380)	(12.960)	(43.220)
Observations	1812	818	712	214	68
Within R-squared	0.794	0.799	0.800	0.771	0.836
Between R-squared	0.000	0.021	0.006	0.044	0.123
Overall R-squared	0.429	0.449	0.423	0.346	0.575

Standard errors in parentheses; * p<0.05, ** p<0.01, *** p<0.001; each column contains the result of an OLS-regression where the reading test scores of a certain group of children is regressed on a measure of ethnic composition of a classroom and several control variables; the first column shows the result for the regression of reading test scores of all non-Western minority children on the percentage of Dutch children; the second to fifth columns show the results for the regression of the reading test scores of Turkish, Moroccan, Surinamese and Antillean children are regressed on the percentage children in a classroom that are not Turkish, Moroccan, Surinamese or Antillean respectively.

Table A.8: Individual fixed-effects regression results for the relationship between ethnic composition measures and math test scores after elimination of outliers

	Math test scores				
	All	Turkish	Moroccan	Surinamese	Antillean
Percentage Dutch	-0.130**				
	(0.040)				
Percentage not Turkish		-0.013			
		(0.097)			
Percentage not Moroccan			0.015		
			(0.085)		
Percentage not Surinamese				-0.068	
				(0.206)	
Percentage not Antillean					1.806***
					(0.449)
Grade	41.440***	41.410***	42.100***	41.210***	35.100***
	(.569)	(1.027)	(0.807)	(1.259)	(2.708)
Weight factor	-.03421	-1.959	1.866	0.310	-5.850
	(1.124)	(1.874)	(1.847)	(3.707)	(3.367)
Version	-13.960***	-15.190***	-16.220***	-10.970***	-15.120***
	(1.217)	(2.136)	(1.861)	(3.367)	(2.917)
School score	-3.828***	-6.566***	-2.735	0	0
	(0.934)	(0.983)	(1.266)	(.)	(.)
Constant	73.170***	97.720***	59.180***	26.120	-126.700***
	(8.816)	(10.560)	(14.250)	(19.300)	(42.070)
Observations	1784	794	640	148	62
Within R-squared	0.881	0.850	0.903	0.943	0.961
Between R-squared	0.032	0.010	0.015	0.055	0.006
Overall R-squared	0.578	0.364	0.554	0.738	0.578

Standard errors in parentheses; * p<0.05, ** p<0.01, *** p<0.001; each column contains the result of an OLS-regression where the math test scores of a certain group of children is regressed on a measure of ethnic composition of a classroom and several control variables; the first column shows the result for the regression of math test scores of all non-Western minority children on the percentage of Dutch children; the second to fifth columns show the results for the regression of the math test scores of Turkish, Moroccan, Surinamese and Antillean children are regressed on the percentage children in a classroom that are not Turkish, Moroccan, Surinamese or Antillean respectively.

Table A.9: Individual fixed-effects regression results for the relationship between ethnic composition measures and reading test scores after elimination of outliers

	Reading test scores				
	All	Turkish	Moroccan	Surinamese	Antillean
Percentage Dutch	-0.090				
	(0.047)				
Percentage not Turkish		-0.167			
		(0.096)			
Percentage not Moroccan			-0.032		
			(0.089)		
Percentage not Surinamese				0.346	
				(0.281)	
Percentage not Antillean					1.321
					(0.732)
Grade	26.520***	26.950***	26.490***	24.030***	28.150***
	(0.486)	(0.749)	(0.778)	(1.875)	(2.915)
Weight factor	-0.871	-0.502	-3.866*	-5.065	4.390
	(1.041)	(1.555)	(1.864)	(4.010)	(5.467)
Version	-1.769*	-2.332*	-0.893	-1.204	-1.046
	(0.756)	(1.119)	(1.291)	(2.460)	(2.230)
School score	0.796	-0.317	2.241	0	0
	(1.482)	(1.366)	(2.361)	(.)	(.)
Constant	-10.370	7.602	-21.170	-28.820	-24.138
	(13.310)	(13.080)	(22.520)	(22.700)	(67.940)
Observations	1784	794	640	148	62
Within R-squared	0.793	0.794	0.796	0.800	0.856
Between R-squared	0.001	0.014	0.003	0.054	0.126
Overall R-squared	0.431	0.454	0.418	0.506	0.598

Standard errors in parentheses; * p<0.05, ** p<0.01, *** p<0.001; each column contains the result of an OLS-regression where the reading test scores of a certain group of children is regressed on a measure of ethnic composition of a classroom and several control variables; the first column shows the result for the regression of reading test scores of all non-Western minority children on the percentage of Dutch children; the second to fifth columns show the results for the regression of the reading test scores of Turkish, Moroccan, Surinamese and Antillean children are regressed on the percentage children in a classroom that are not Turkish, Moroccan, Surinamese or Antillean respectively.

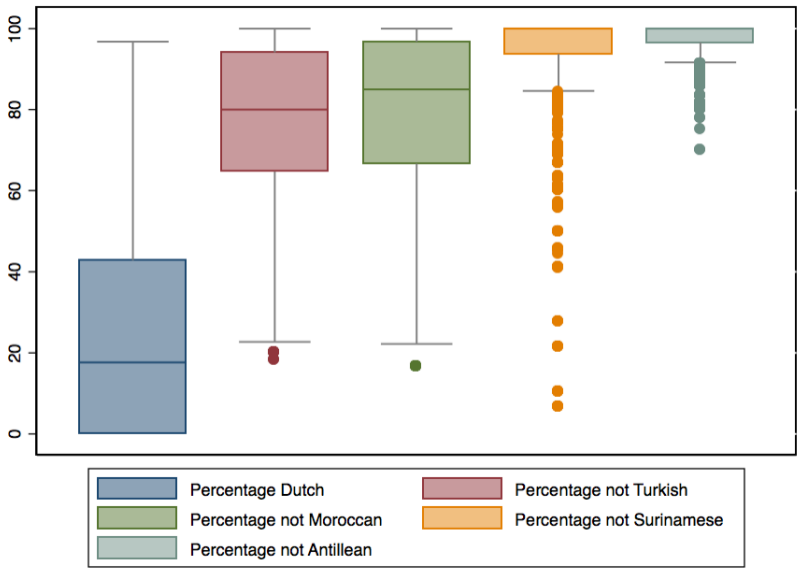

Figure A.3: Box-plots of measures of ethnic classroom composition

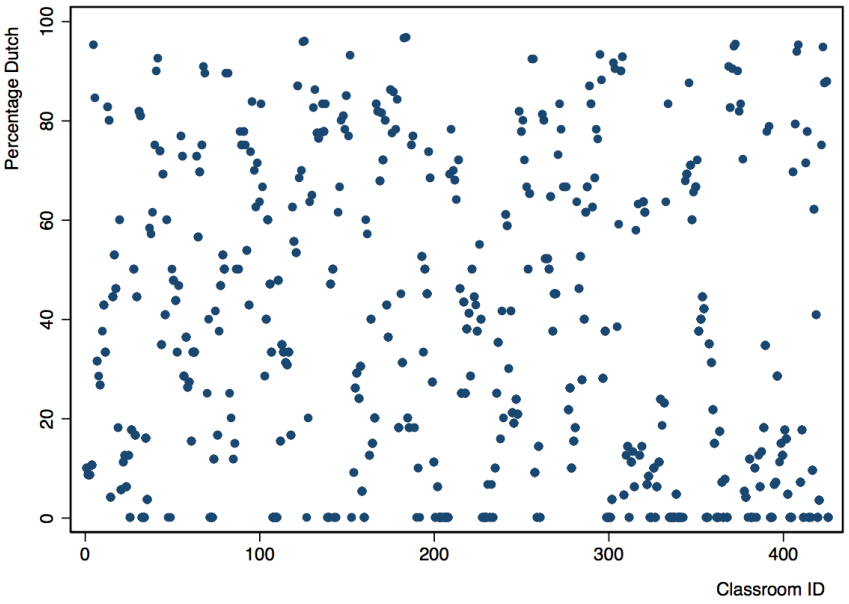

Figure A.4: Percentage of children that are Dutch of each individual classroom

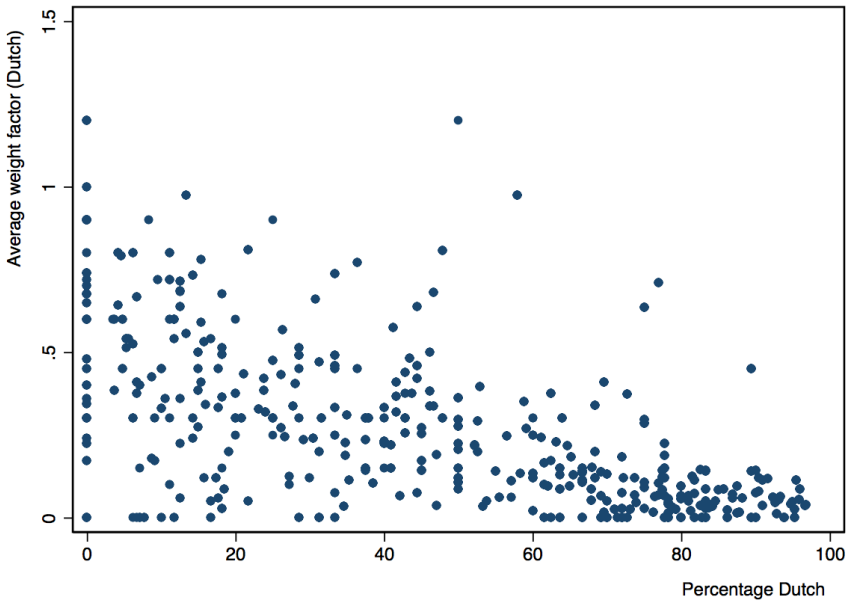

Figure A.5: Average weight factor of children with a Dutch background in a classroom in terms of percentage of children with a Dutch background