

# **Peer Ethnicity and Achievement: a Meta-analysis Into the Compositional Effect**

## **Abstract**

This study reports a meta-analysis on the effects of ethnic minority share in school on achievement test scores. Best evidence from the studies that have appeared thus far on this topic shows that these compositional effects appear small in general, but may be larger when the ethnic minority group is African Americans in the USA, than when the minority group consists of immigrants. A high share of students from an ethnic minority group seems to affect the achievement from students belonging to the same ethnic group more, than the achievement of students belonging to the ethnic majority or to other ethnic minority groups. Effects of the share of immigrants on test scores of ethnic majority students even seem to be close to zero. Several robustness checks confirm our results. The review concludes with a discussion of implications for research and policy practice.

**Keywords:** academic achievement, meta-analysis, racial composition, school segregation, ethnic groups

## **Introduction**

In many countries around the world, there is a strong imbalance in the numbers of students from ethnic minorities in different schools. According to Rumberger & Palardy (2004), in the United States, over 70% of African American and Hispanic students attend schools where ethnic minorities constitute the majority of the population. In the Netherlands, ethnic minority children attend elementary schools with on average 70% ethnic minority students, while this number is 10% for ethnic majority students (Gijsberts, 2003); in the United Kingdom, two-third of all ethnic minority students would have to change schools in order to obtain an equal distribution (Burgess & Wilson, 2005).

A common viewpoint is that the existence of schools with high proportions of ethnic minority students leads to negative educational and social consequences. Consequently, in the last decades, policy makers all over the world have developed strategies to cope with segregation, ranging from busing, strategic redrawing of attendance zones, magnet schools to setting quota for admissions.. The debate on the nature and effectiveness of these strategies among policy makers and researchers continues till today, as the recent United States Supreme Court Decision in the Seattle and Louisville cases shows (Meredith, 2007; Parents Involved in Community Schools, 2007).

In order to design evidence-based policy strategies to deal with school segregation effectively, systematic knowledge about the extent to which the share of ethnic minorities in classes and schools affects students' educational achievement is essential. Therefore, it is important to reconcile the findings from previous studies on the effects of minority share on the educational achievement of students by means of a meta-analysis, Although several scholars have conducted studies on the compositional

effect related to ethnicity, such a meta-analysis has not been carried out yet. The present study aimed to fill this gap.

We will first describe the many similarities in the situations of ethnic minorities around the world, in the segregation they face, and in the associated compositional effects. Next, we show that despite these similarities, there are also differences between ethnic groups that we have to take into account when pooling together the previous research on the effects of composition. After that, we describe our meta-analytic method, and in the next section the results. The implications and limitations of our findings are discussed in the final section.

### ***Ethnic Minority Share in Schools: Similarities and Effects***

The effect of ethnic minority share on academic achievement has been described as part of a more general “compositional effect”. A compositional effect is the effect that going to school with children with certain background characteristics has on specified outcomes. Background characteristics that have been studied include socioeconomic status (SES) (e.g. Lee & Bryk, 1989; Rivkin, 2001; Van Ewijk & Slegers, 2007; Willms, 1986), ability (e.g. Sacerdote, 2001; Vigdor & Nechyba, 2004; Zimmerman, 2003), sex (e.g. De Fraine et al., 2003; Hutchison, 2003), and, the focus of this study, ethnicity. The outcome variable that has been studied the most often is academic achievement. This is also the outcome we concentrate on.

The effect of ethnic minority share in schools and classes on academic achievement has been described using various terms. Most studies describe it as “compositional effect” or simply as the effect of composition. A few others speak about “contextual” or “school-mix” effect, but no fundamental difference exists between the phenomena these terms refer to (Harker & Tymms, 2004). Consequently, some studies even use several terms alternatingly (Harker & Nash, 1996; Peetsma,

Van der Veen, Koopman, & Van Schooten, 2005). Economists ordinarily use the term “peer effect”. With this, they refer to the same phenomenon, although generally they place more emphasis on measuring the “pure” effect of minority share, filtered from all correlates and potential biases. In this study, we will use the terms compositional effects and effects of composition, since they are the most widely used in the Social Sciences.

The compositional effect of ethnic minority share has been studied for several different minorities in several different countries. In virtually all of these countries, there is a substantial amount of school segregation along ethnic lines (Burgess & Wilson, 2005; Gijssberts, 2003; Gorard & Smith, 2004; Lauder & Hughes, 1999; McEwan, 2004; Rumberger & Palardy, 2004; Schindler-Rangvid, 2007). Ethnic minorities are also similar in many challenges they face.<sup>1</sup> In the USA, large achievement gaps exist between Afro Americans and Whites (Thernstrom & Thernstrom, 2003), and between Hispanics and Whites (Clotfelter, Ladd, & Vigdor, 2006b). This pattern of underachievement of ethnic minorities in comparison with the country’s ethnic majority can be seen in most countries of the Organisation for Economic Co-Operation and Development (OECD) (Demack, Drew & Grimsley, 2000; OECD, 2003; Schindler-Rangvid, 2007). Outside of the OECD, McEwan (2004), shows that students from the indigenous population in Bolivia score lower than non-indigenous students. A similar indigenous – non-indigenous gap is found by Hoxby (2000) for Native Americans in Texas.

Not only in school achievement gaps, but also in a variety of other situations, ethnic minorities in different countries face similar gaps and problems. Wage-gaps, ethnic employment discrimination, racism and discrimination against ethnic

minorities have been shown to exist in countries all over the world (Grogger, 1996; Riach & Rich, 2002; Canessa, 2004; Stevens, 2007).

In order to cope with the large achievement gaps and other problems faced by ethnic minorities, politicians in different countries have developed strategies to combat school segregation along ethnic lines. Many strategies, including residential solutions such as mixing of the housing stock and vouchers to help people move to better neighborhoods, strategic redrawing of attendance zones, busing and magnet schools to attract ethnic majority students to schools with a high ethnic minorities share, have often had only limited effects (Angrist & Lang, 2004; Burgess & Wilson, 2005; Driessen, 2002; Katz et al., 2001; Musterd, 2005; Schindler-Rangvid, 2007). Some potentially effective measures accept segregation as a given and aim to improve the quality of high ethnic minority share schools by increasing funding to those schools, or by increasing the salaries of teachers to keep talented teachers in such schools (Clotfelter, Ladd & Vigdor, 2006a; Driessen, 2002). Possibly effective for combating school segregation itself, are quota for admissions. If these quota explicitly refer to ethnicity, however, legal barriers may form an obstacle, see the recent US Supreme Court decision in the Seattle and Louisville cases (Meredith, 2007; Parents Involved in Community Schools, 2007) and similar issues raised in The Netherlands by the Dutch Equal Treatment Commission, an important anti-discriminatory advisory board.

#### *Differences Between Ethnic Minority Groups: Differential Effects*

Although, in broad lines, the situations of disadvantaged ethnic minorities are comparable across countries, there are also some important differences. Of course, the wideness of (achievement) gaps and the extent of segregation and discrimination, vary. But there are also some more fundamental differences. In this, we can roughly

make a division into three ethnic minority groups according to their history and situation: immigrants, African Americans and indigenous people. Immigrants differ from the other two groups in that they chose to move to the host country because of advantages it had to offer to them: better economic perspectives, freedom from oppression, the presence of family members in the new country, or a combination of these. Immigrants in different countries are similar to each other in their adaptation problems to the culture of the host country and (usually) in their unfamiliarity with the host country's main languages<sup>2</sup>. African Americans stand out through the forced way in which their ancestors had to immigrate and through a long history of overt and legal discrimination. Also, the achievement gap with the ethnic majority is larger for African Americans than for most other groups (Thernstrom & Thernstrom, 2003). Under "indigenous people", we group the original inhabitants of countries once colonized by Europeans and now dominated by their descendants. Native Americans, the Maori of New Zealand and the indigenous people of Latin America share a history of marginalization and oppression by the still dominant population group and often of denial of their own culture: no education in their own language, explicit or implicit discouragement of the practicing of their old cultural habits, et cetera.

Despite these historical and situational differences between groups of ethnic minorities, scholars have not made a theoretical distinction between compositional effects associated with each of the groups. When authors, in one and the same study, examine compositional effects related to more than one ethnic minority, they just compare effect sizes, assuming that the mechanisms underlying these effects are the same for each of the ethnic minority groups (e.g. Hanushek, Kain, & Rivkin, 2002; Harker & Nash, 1996; Hoxby, 2000; Vigdor & Nechyba, 2004). We follow this convention and do not treat effects related to different groups as theoretically

different. We do, however, think that there may be differential effects of composition. Compositional effects may for example be stronger when the ethnic minority group consists of Afro Americans than when the minority consists of immigrants. And ethnic minority students themselves may or may not be affected more strongly than students belonging to the ethnic majority (Angrist & Lang, 2004; Caldas & Bankston, 1998; Gould, Lavy, & Paserman, 2004; Hanushek et al., 2002; Hoxby, 2000).

By performing a meta-analysis, we try to bring structure into the wide variety of results that have been found so far on the compositional effect. An important goal of this meta-analysis is to find out to what extent compositional effects differ between ethnic groups. We therefore will focus on both the ethnic group that causes the effect and the group that is affected: both may make an important difference for the size of the compositional effect. By introducing a classification in ethnic groups, we can do justice to differences that may exist between ethnic groups, while at the same time it enables us to make generalizations about the subject at large. Nevertheless, we do realize that, when we, for example, group together all immigrants that are in a disadvantaged situation, we will not be able to capture possible differences between students within one specific ethnic minority group.

Furthermore, we will also investigate whether other characteristics of the included studies, such as their samples (student age, test type) and characteristics of their estimation models affect their effect sizes. In this way, we try to shed light on why the reported effects of increasing the minority share range from strongly negative effects (e.g. Bankston & Caldas, 1996; Strand, 1997), via small effects (e.g. Angrist & Lang, 2004) and no significant effects (e.g. Rivkin, 2000), to even some estimates that suggest a positive effect of ethnic minority share on test scores (e.g. Link & Mulligan, 1991; Vigdor & Nechyba).<sup>3</sup>

### *Causes of the Compositional Effect*

Many causes for the compositional effect of ethnicity have been proposed, although virtually no study quantitatively examines the link between proposed causes and the taking place of a compositional effect. In the great majority of studies, the size of the effect is estimated and some causes are listed as potential explanations for the effect. Nevertheless, for a good understanding of compositional effects, it is essential to understand how they may work and in which situations each cause might be of importance. The proposed causes why a student would perform poorer as the ethnic minority share increases, can roughly be grouped into four categories: direct peer interaction, teacher practice, school quality and research artifacts.

#### *1. Direct Peer Interaction*

Students who go to school together and interact on a daily basis will inevitably influence each others' attitudes and behavior and may thus influence each other's school performance. Poorly motivated students may convince others that it is not worth doing your best at school, or even put pressure on others not to excel (Driessen, 2002; Hanushek et al., 2002; Harker & Tymms, 2004). Disruptive students may prevent others from learning, while students with more knowledge may help their classmates. Peers may also influence students' learning through the general level of conversation and through the out-of-class activities that children choose to do (Harker & Tymms, 2004). Most of the ways in which ethnic composition affects achievement are basically unrelated to students' ethnicity per se. Peers' ethnicity is then mainly relevant because of its correlation with variables such as motivation, socioeconomic status, and ability. Because it is virtually impossible to separate effects of ethnicity

from effects of its correlates, researchers are generally interested in the effects of peer ethnicity including all its correlates (Card & Rothstein, 2006).

Two channels of direct peer interaction that are inseparably linked to ethnicity are: 1. tensions between races that may interfere with learning (Hoxby, 2000) and 2. differences between the ethnic minority students' mother tongue and the country's official language. The higher the minority share, the higher the chance that minority students will mainly speak their mother tongue among themselves in school, have less contact with the majority's language and will therefore learn the official language less well. (Driessen, 2002; Peetsma et al., 2005). This factor is expected to be most important for immigrants and then especially for their language performance.

## *2. Teacher practice and teacher quality*

Teachers may adjust their teaching style to the group of children in the class (Harker & Tymms, 2004; Hoxby, 2000; Thrupp, 1995). This may be a deliberate choice, for example if teachers adapt their teaching style to the specific needs of their students. This may positively affect the students with low skill-levels and specific needs, but the achievement of higher-skilled students without these specific needs may suffer: they would have performed better in a class with a different composition. Teachers can also have unjustified low expectations of ethnic minority students that may (often unintentionally) be communicated to the entire class. These low expectations may affect students' beliefs in their own competence and subsequently lower their performances. This phenomenon is known by the name of "Pygmalion effect" (Rosenthal & Jacobson, 1968).

Furthermore, ethnic minority share may also be related to the quality of the teachers and the teaching staff. In many countries, schools with a high minority share,

or more generally, with a less privileged intake, have difficulties in attracting highly qualified teachers (OECD, 2001; Scafidi, Sjoquist, & Stinebrickner, 2007). As a result, they often end up with less-qualified or less experienced teachers and have to struggle with higher teacher-turnover rates (Clotfelter, et al., 2006a).

### *3. Artifacts*

A final reason why compositional effects may be found is that they only appear as artifacts of the statistical analysis used (Harker & Tymms, 2004; Hauser, 1970; Evans, Oates & Schwab, 1992; Hanushek, Kain, Markman & Rivkin, 2001; Nash, 2003). Even without a causal effect of composition, students in a school or class with a high proportion of ethnic minorities will generally score lower on achievement tests. This is because there are student characteristics that simultaneously increase the expected share of ethnic minority students in their school or class and that negatively affect their achievement. For example, an ethnic majority child who goes to school with many ethnic minority children and gets low test scores, may either perform poorly because of a compositional effect, or just because his parents are poor and therefore live in a poor neighborhood where the ethnic minority share is high and where the schools also have a high ethnic minority share. Effects of ethnic minority share and of background can easily be confounded here. Such a statistical artifact is often denoted as “omitted variables bias”, since including all variables as covariates that are simultaneously correlated with proportion ethnic minority students and with achievement will solve the problem. It is, however, impossible to prove that all such covariates have been included. This is especially because the processes that may cause a child’s ability and his/her school composition to be correlated can be very subtle. Take for example parents who rightfully have high expectations of their child

and therefore try harder to get their child into a school with a good reputation (and a low ethnic minority share). This may cause a confounding of effects of a child's own talent and those of school composition. But the omitted variables that are relevant here, true talent and support at home (stimulation, support with homework, etc.), are very hard to measure in a reliable way.

For such reasons, Economists have developed statistical techniques that can, under certain assumptions, overcome this problem and yield unbiased estimates. Instrumental variables (IV) approaches and fixed effects analyses are the most important of these techniques here. It is often argued that dealing with omitted variables is essential when estimating compositional effects (e.g. Harker & Tymms, 2004; Evans et al., 1992), because researchers otherwise risk overestimating them.

## **Method**

### ***Criteria for Inclusion***

In this study, we review studies on the effect of ethnic minority share on students' scores on achievement tests. To be included in this review, a study has to fulfill the following criteria:

1. It gives estimates of the effect of an increase in the proportion of ethnic minority students. Studies defining composition only by rough categories (for instance schools with more versus less than a certain percentage minority students) and studies comparing segregated with desegregated schools were excluded.
2. The ethnic minority in question scores lower on achievement tests than the country's ethnic majority. Estimates from the United States, for example, on effects of the share of Asians (e.g. Hoxby, 2000) are excluded since (East-

)Asians on average do not score lower than ethnic majority students, nor are they in a disadvantaged situation as described before, nor do most of the listed potential causes of the compositional effect apply.

3. The dependent variable is individual students' scores on tests of mathematics, language, science or general academic achievement (being combinations of the three other types of tests). This excludes rough categories such as whether students drop out or pass exams.
4. The estimation (regression) model used in the study, includes individual ethnicity as a covariate. Individual ethnicity and proportion ethnic minorities among peers are highly correlated, so that not including individual ethnicity would lead the proportion minority variable to serve as a proxy for individual students' ethnicity. This would inevitably lead to a considerable overestimation of the compositional effect.
5. The students in the sample are in primary or secondary (high) school (6-18 years old).
6. The study is published or presented no earlier than January 1986 and before January 2006. Note that while the aim of this study is to cover this twenty-year period, the earliest included studies appeared in 1996. Partly this lack of earlier studies can be explained by an increased attention to the topic in later years (up to the 1980s, focus in the USA was more on effects of desegregation attempts than on pure compositional effects, while elsewhere, research into compositional effects focused more on socioeconomic status composition); partly it can be explained by the introduction and dissemination of statistical techniques and programs that made it easier to conduct appropriate analyses on large databases.

7. The study is written in English.
8. The study uses students' present ("level") test scores, not gain scores, as the dependent variable in its model. Estimates using gain scores refer to a different type of effect than estimates using level scores, and both types of estimates cannot be compared quantitatively, so that they would require separate meta-analyses. From the studies fulfilling the previous criteria, however, all but one (Hanushek et al., 2002), used level equations. Hence, our analyses will remain constrained to estimates from level equations.

### ***Identification of Studies***

Both published and unpublished studies were included in this meta-analysis. Eligible studies were identified by systematic searches of electronic databases related to different scientific disciplines including ERIC, Sociological Abstracts and EconLit. Search terms included combinations of the terms (racial) composition, compositional effect, contextual effect, peer, peer effect, peer influence, racial factors, ethnic groups, racial segregation, classroom environment and achievement. We thoroughly examined each of the identified studies for references to additional studies. In this way, several more studies on the compositional effect were identified.

### ***Coding Procedure***

Each study was coded by one of the researchers using a formal scheme. To obtain a high degree of coding reliability, all codings were independently verified by the other researcher. Differences in codings were discussed until consensus was reached. Using the coding form, effect sizes with their standard errors and information on relevant potential moderators of the compositional effect were systematically recorded.

Potential moderators will be discussed in detail below. Whenever information necessary for coding was not reported in the study, we contacted the author(s). A few studies had to be excluded, because the authors could not be contacted, or because they were unable to retrieve information that was essential for inclusion.

Most studies gave several estimates of compositional effects. In some cases, a clear preference was given for estimates from one (final) model and other estimates were only shown as intermediate steps to arrive at this final model. If this was the case, only estimates from the preferred model were included; if no clear preference was given, all estimates were included. The final database included 177 estimates from 13 studies (see Table 1).

-----  
Insert Table 1 about here  
-----

***Calculation and Weighting of Effect Sizes***

All included studies used some form of regression analysis. We “standardized” each reported effect estimate (being a regression coefficient) into a parameter that reflects the effect of an increase in the proportion ethnic minority students on standardized test scores. (That is: reported parameter estimates were divided by the standard deviation of the achievement test used in the study; if the parameter estimate for example referred to an increase in the minority share by one percent, the estimate was multiplied by 100 to make it refer to an increase in proportion.) The standard errors to the estimates, which were necessary for the weighting procedure described below, underwent the same linear transformation.<sup>4</sup>

We estimate a mixed-effects meta-regression (Overton, 1998; Lipsey & Wilson, 2001) of the following form:

$$(1) \quad T_{ij} = \beta_0 + \sum_{k=1}^l \beta_k X_{kij} + e_{ij} + u_{ij}$$

In this,  $T_{ij}$  is the  $i$ 'th effect estimate reported in study  $j$  of the “true” compositional effect. Because of sampling error,  $T_{ij}$  can deviate from the true value of the compositional effect, which is reflected in the sampling error term,  $e_{ij}$ , which is the standard error to the estimate as reported in the study. Both  $T_{ij}$  and  $e_{ij}$  have been standardized as described above. We assume that the true effect is not constant over all estimates, and that this variation can be captured partially, but not completely, by our set of moderators,  $X_k$ . Therefore, we estimate a mixed-effects meta-regression. This means that we see the estimates included in this analysis as a random sample from all potential estimates that could have been made of the effect. The true effect size is assumedly normally distributed over all potential studies, with variance  $\sigma_\theta^2$ . To capture this variation, we add an additional error term,  $u_{ij}$ , to our model (Overton, 1998; Raudenbush, 1994). Adding this term enables us to generalize beyond the particular set of estimates we included (Hedges & Vevea, 1998).

Each estimate in the meta-regression is weighted by the inverse of its total variance (Lipsey & Wilson, 2001; Raudenbush, 1994):

$$(2) \quad w_{ij} = \frac{1}{v_{ij} + \sigma_\theta^2},$$

in which  $v_{ij}$  is the squared of the sampling error,  $e_{ij}$ . As discussed, most studies contributed several effect estimates for which some of the moderators,  $X_k$ , are usually different. Weighting simply by  $w_{ij}$  as described, means that we ignore dependencies between estimates coming from the same study. Studies contributing several estimates that are not completely independent of each other, therefore get a disproportionately large weight. We adjust our weighting procedure to take this into account properly, as is described in Appendix 1. We estimate the meta-regression from Equation (1) with

these adjusted weights using SPSS by means of restricted maximum likelihood (Hox, 2002; Thompson & Higgins, 2002).

Note that although a few studies contribute a large number of estimates, the results are by no means predominantly determined by these few studies: the regression parameters we estimate are determined by *variation* between contributed effect estimates, and this often comes from variation between studies, so that the studies contributing only one or two estimates are very important as well. Also, because of the adjustment described in Appendix 1, individual estimates from the few studies that contribute many estimates, receive lower weights than individual estimates from the average other study. In Appendix 2, we show that deleting the two studies that contribute the highest number of estimates, does not affect the robustness of our results in an important way.

In the basic meta-regressions following Equation (1), we start with an empty model, and in three following models each time add a set of covariates,  $X_k$ . After this, we additionally conduct study fixed effects analyses in order to check the robustness of our results. In these analyses, we combine meta-regression with panel fixed effects techniques as will be described next.

### ***Study Fixed Effects Analyses***

In meta-analyses (as in most analyses on non-experimental data), there is a risk that omitted variables may bias the results. Omitted variables are variables that are simultaneously correlated with the dependent variable (here: the effect sizes), and with the independent variable (here: characteristics of included studies). If these omitted variables are not included as covariates in a regression, they will bias the results. To give an example: in our meta-analysis, we compare effects of immigrant

share with effects of Afro American share. Suppose that the included studies which studied the effects of immigrant share on student achievement do not correct for prior achievement very often. This will lead to an overestimation of the effects of immigrant share (see also the paragraph on moderators below). If we, in our model, do not correct for the study characteristic “corrected for prior achievement”, we may falsely conclude that the effect of immigrant share is larger than the effect of African American share. But this may only be an artifact, arising because of our failure to correct for an important covariate. In our meta-analysis, we actually do correct for this particular covariate. But we cannot be sure that every other relevant covariate is included, and therefore we potentially run the risk of omitted variables bias.

Using a meta-regression, instead of a meta-ANOVA, greatly reduces this risk, because at least some covariates can be included. But it does not completely solve it: included studies differ on many characteristics and it is impossible to include a covariate for each of those. We, therefore, check the robustness of the results against omitted variables bias by combining fixed effects panel data techniques with meta-regression analysis. In a fixed effects meta-regression, we use the fact that most of the included studies contributed several effect estimates. We filter out all systematic differences in effect sizes *between* studies. All identification now comes from differences *within* studies in the characteristics of the effect estimates they contributed. By taking out the between-study differences, we make sure that study characteristics that are correlated to our covariates cannot bias our results. We estimate a meta-regression of the following form:

$$(3) \quad T_{ij} = \alpha_j + \sum_{k=1}^l \beta_k X_{kij} + e_{ij}$$

In this,  $\alpha_j$  stands for a fixed effect (cf. dummy variable) per study; the other terms are similar to those in Equation (1).<sup>5,6</sup> Note that using this specification, no estimates on characteristics that are constant within each of the studies can be obtained. Also, no constant,  $\beta_0$ , can be estimated. And because we use a dummy variable per study,  $R^2$  is artificially inflated and hence loses its meaning. Since not all studies contribute more than one estimate, the number of included estimates is lower than in the previous analyses: 174 estimates from 10 studies were included (see Table 1).

### ***Moderators***

The main question in this meta-analysis is whether the size of the compositional effect differs between the ethnic groups that we distinguish. Hence, the main covariates in each analysis are dummies indicating whether an effect pertained to increasing the share of African Americans, immigrants, or indigenous children and dummies indicating whether effects were on test scores of students from the ethnic majority, on students from an ethnic minority, or on the entire student population.

We also investigate whether studies' sample and estimation model characteristics affect the size of the compositional effects they found. Two sample characteristics that might moderate the compositional effects are age and test type. We expect that as children get older, the influence of adults such as parents and teachers on their behavior may decrease, while the influence of peers of their own age increases. Hence, the compositional effect may increase in size as students get older. The main hypothesis on why compositional effects would differ between test types is the so-called contact-hypothesis: a high concentration of ethnic minority students who speak another language at home than the language of instruction at school, may lead

to less contact with, and hence, proficiency in, the instructional language (Driessen, 2002; Peetsma et al., 2005). In the case of immigrants, this would suggest a stronger compositional effect on language test scores.

Differences between the studies included in this meta-analysis in how they modeled and analyzed the effects of composition lead to a set of five potential moderators. The most fundamental difference in approaches is that between studies that name the effect peer effect and those that write about compositional or contextual effects. The term peer effects is mainly preferred and used by Economists, who have a relatively restrictive interpretation of the effect: it is usually meant to be cleaned from all correlates that are not by definition tied to minority share, such as school and teacher characteristics. As a consequence, Economists measuring peer effects generally have a stronger focus on avoiding statistical artifacts such as omitted variables bias and may therefore find weaker effects.

A second characteristic of studies is the level at which they measured minority share: at the level of the class, or at the broader level of the entire cohort or school. Of the proposed causes of the compositional effect, especially those causes related to direct peer interaction and those related to the teacher in the class are dependent on the composition of the class. Composition of cohort or school is only relevant here, to the extent that it is an approximation of the composition of the class. Since approximations contain noise, estimates from data in which composition is measured at cohort or school level may suffer from attenuation bias, i.e. a bias toward zero.<sup>7</sup> In so far that compositional effects are caused by school quality differences, as several authors propose, it makes no difference whether composition is measured at class-level or at school-level.

The three other study characteristics pertain to the covariates studies used in their models: socioeconomic status, ability and students' prior achievement. Ethnicity correlates with socioeconomic status and ability. Therefore, studies that correct for average socioeconomic status, may partially explain away the effect of ethnicity, and hence find smaller effects. Similarly, studies that correct for average ability level (often measured by averages on a prior test), may also partially explain away the effect. If a study corrects for individual students' prior achievement, this may affect its estimates for a different reason: estimating compositional effects without correcting for individual students' prior achievement or ability, is very likely to lead to overestimation (Goldhaber & Brewer, 1997; Hanushek et al., 2002; Ho Sui Chu & Willms, 1996; Rumberger & Palardy, 2005). The first reason for this is that prior achievement may have influenced which school, class, or track the student currently attends, and thus also the current ethnic minority share among his peers (in lower tracks, there may be more ethnic minority students). Researchers who do not correct for individual prior achievement may therefore confound effects of prior achievement with effects of current minority share. The second reason is that prior achievement may be affected by prior composition. Not correcting for prior achievement then leads to a confounding of effects of current composition and those of composition in the past. As composition in the past is strongly correlated to current composition, the estimated effect will be an accumulation of the compositional effects that a student has experienced over all previous school years.

## Results

### *Meta-regression: general differential effects*

Table 2 shows the results from the first basic set of meta-regressions. After estimating an empty model, we add dummies that indicate the ethnic group that was studied. Next, we add other characteristics of the samples used in the included studies, and finally, we add characteristics of their estimation models.

-----  
Insert Table 2 about here  
-----

The empty model (model 1) gives the weighted average effect size over all included studies. Recall that effect sizes refer to increasing a proportion: on average, the studies find an effect size of -0.18, which indicates that increasing the share of students from an ethnic minority by ten percentage points is associated with a decrease in individual students' test scores by 0.018 standard deviation. There is considerable variation between the included studies in the effects they find, as is indicated by the highly significant systematic variance component. By adding covariates in the following models, the systematic variance will decrease.

Model 2 adds dummies indicating the ethnic groups to which the effects are related. The model should be read additively: the constant refers to the effect of African American share (the omitted category in the first set of dummies) on test scores of students from this same ethnic minority group (the omitted category in the second set of dummies). For the effect of African American share on test scores of students from the ethnic majority, add 0.128 to the constant; for effects of immigrants on immigrants of the same ethnicity, add 0.161 to the constant, and for effects of immigrants on students from another ethnicity, add 0.161 plus 0.180, and so forth.

Model 3 adds other sample characteristics. The covariate age is centered at 12, and the omitted category for test is math<sup>8</sup>, so that the constant now obtains the interpretation of the effect on math scores of 12-year olds. In the final model (model 4), where we add other characteristics of models and studies, the clear interpretation of the constant and the additive nature of the coefficients remain, since for each (dummy) covariate, the omitted category refers to the arguably best way of measuring compositional effects.<sup>9</sup>

The size of the compositional effect seems to differ substantially across the ethnic groups. Increasing the share of African Americans is associated with a considerably stronger negative effect on test scores than increasing the share of immigrants. Effects related to the share of indigenous students lie close to those of (and cannot statistically be distinguished from) the African American share. Increasing ethnic minority shares mainly seems to affect students who belong to the same ethnic minority groups. Students from the ethnic majority and students from other ethnic minorities appear to be affected much less (as is indicated by the positive parameter estimates). Effects pertaining to test scores of the entire population, being a mix of different ethnic minority groups and the ethnic majority, should lie somewhere in between of those, depending on the exact ratios of the different groups. That the point estimate lies higher than that for the ethnic majority may be due to sampling variance.

By adding up the coefficients as described, we see that the compositional effects seem small: increasing the Afro American or indigenous students share by ten percentage points is associated with a decrease in test scores of students from those same minorities by about 0.027 standard deviation. The effect of this on the performance of other ethnic groups (either minority or majority) is about half that

size. And so is the effect of immigrant share on immigrants' own test scores. The immigrant share seems to have an effect of around zero on test scores of students from other ethnic groups and on students from the ethnic majority.

Table 2 also shows that age nor test type moderate the size of the compositional effect found. The interaction of language test x effects of immigrants, which was expected to be negative, is far from significant. This is not surprising, given the small effects of immigrant share in general. Researchers focusing on "peer", instead of on "compositional" or "contextual" effects, do not find considerably smaller effects, despite their somewhat more delimited definition of the effect and their stronger focus on avoiding bias due to omitted variables.<sup>10</sup>

The point estimate for the covariate indicating whether a correction for individual students' prior achievement was made is quite strong. This suggests that failing to do so leads to a substantial overestimation of the effects of composition. But it has such a large standard error that it is not significant. This might either mean that correcting for prior achievement is not necessary, or that it is necessary only in certain instances, i.e. there where the compositional effect is substantial: if a compositional effect is small (for instance the effect of immigrant share on majority's test scores), it makes no difference for the estimate if (due to omitting pre scores) it picks up effects of composition in the past.<sup>11</sup>

The small coefficient for the effect of correcting for average SES suggests that the lower average SES of ethnic minorities at best only explains a small part of the compositional effect. This might be an underestimation, as average SES has been shown to have a considerable effect on test scores (e.g. Opdenakker et al., 2002; McEwan, 2003; Willms, 1986) and given the much lower average SES of ethnic minorities. This underestimation may occur because of the way most studies

measured average SES. As Van Ewijk & Slegers (2007) show, using a rough, unreliable measure, a measure that only captures certain parts of the SES-concept, or measuring average SES at cohort- or school-level instead of at class-level, leads to a bias towards zero in the coefficient to average SES. Correcting for average SES in these cases affects the size of the ethnicity compositional effect to a much lesser extent. Again, the small effect may also be a result of the general small size of the compositional effect, which means that correcting for average SES cannot make a large difference in absolute sense by definition. The same applies for the level at which ethnic minority share is measured: we do not find the expected difference in effect size between those instances where composition was measured at school / cohort level and those where it was measured at class level. The near-zero coefficient for the effect of correcting for average ability level suggests that ability is also not a main channel through which composition according to ethnicity has an effect.

The variance explained in the final model is relatively low, with a bit above 16%. Nevertheless, the *systematic* variance component has shrunk by about half by adding the set of predictors. This means that our model does quite well on explaining that part of the variance that can be explained: much of the unexplained variance is due to sampling error which can never be explained by adding covariates. That this sampling variance is relatively so large is due to the small sizes of the compositional effects: in some cases it is about zero. The closer effects are to zero, the larger the part of the variance between estimates that will be due to noise / sampling variance, which can by definition not be related to any possible covariate. In the most extreme case, in which all studies estimate the same effect with a true size of zero, all differences between these estimates will be noise and no variance can be explained.

***Meta-regression: differential effects specified***

In the first set of models, we made the implicit assumption that e.g. the coefficient for “effect is of immigrant share” would be the same in size, irrespective if the affected group was the ethnic majority, or ethnic minority students. We may, however, have ignored some interaction effects in this way. Our implicit assumption gave us a higher power and sharper point estimates and had the added advantage of an easier interpretation, but it may have been too strong. In the following, alternative model, we try to verify this assumption by using a finer subdivision of the “effects of” and “effects on”. Each combination of “effect of” and “on” will be entered as a separate covariate, with the reference category being the effect of African American share on test scores of African Americans. All other combinations of effect on / of are compared to this reference category which is captured by the constant. We also enter the set of other covariates from the previous model. The results from this analysis are presented in Table 3.

-----

Insert Table 3 about here

-----

-----

Insert Table 4 about here

-----

Effects should again be interpreted additively: the constant plus the parameter to an effect on / of combination gives the estimated effect for that combination. Table 4 shows these added-up estimates and the number of studies and effect estimates that were included on each combination in this meta-analysis. Note that for indigenous students, only effects on the entire student population are estimated, because of a lack of studies giving estimates of the effect of indigenous share on other groups.<sup>12</sup>

The estimates from this second specification generally follow those from the first analysis: effects of African Americans and indigenous seem about equal and appear larger than effects of immigrants. Effects on the own minority group turn out larger than effects on the ethnic majority. Hence, the second analysis strengthens the conclusions we drew from the first analysis. One new interaction that shows up is that the “African American on other minority” (viz. Hispanics) effect seems about equal to the “African American on same minority” (i.e. on African Americans) effect. The “immigrant on other minority effect”, on the contrary, appears much weaker than the “immigrant on same minority effect”, and even (although nonsignificantly) seems positive.<sup>13</sup> As in the first analysis, none of the covariates that are unrelated to the ethnic groups the effect is of or on, is significant. The point estimates for these parameters are also very similar in both analyses.

***Study fixed effects analyses***

Table 5 and 6 show the results from the study fixed effects models. These models use only within-study variation: variation in the characteristics of the various estimates contributed by a study. Hence, some parameters are only estimated using one or two studies and some of the covariates from the previous models could not be included. The study fixed effects models therefore mainly serve as a robustness check. Recall that using this type of model,  $R^2$  loses meaning and that a constant cannot be estimated, so that we cannot give an estimate for the reference categories (the effect of African American share on students from that same minority). However, we can compare the size of all other effects to that for the reference categories.

-----

Insert Table 5 about here

-----

The specification of the first study fixed effects analysis follows the first “ordinary” meta-regression. The pattern of results is about the same: the effects *on* students from the ethnic majority and *on* students from another ethnic minority again seem smaller than effects *on* the same ethnic minority, but the difference is smaller. The effect *of* indigenous share now seems stronger than that *of* African American share, but the standard error is very large, so that like the estimate in the first meta-regression, the difference cannot be statistically distinguished from zero. The effect *of* immigrant share, again, is significantly smaller than the effect *of* African American share.

Only three studies reported estimates on more than one age group. From those studies, it seems that the compositional effect is stronger among older students. That this effect did not show up in the earlier analyses may be because now only within-study differences in the three studies drive the result, while before, unmeasured between-study differences may have obscured the age-effect.

The second study fixed effects analysis follows the second ordinary meta-regression. Again, the parameter estimates generally follow those from the previous analysis, although the exact parameter sizes and a few significance levels differ. Notably, the African American on majority / entire population estimates are weaker and the immigrant on entire population estimate is larger and now significant. Unlike in the first study fixed effects analysis, the effect of indigenous share now seems weaker than that found in the ordinary meta-regression, but is again estimated very imprecisely, so that strong conclusions cannot be attached. Once more, there is a strong indication that compositional effects increase in strength as children get older. So, the two study fixed effects meta-regressions generally confirm the robustness of

our earlier analyses. Some parameters do differ in size from those in the ordinary meta-regressions, but the pattern of results is the same. The notable new finding from the study fixed effects analyses is the significant effect of age.

-----  
Insert Table 6 about here  
-----

**Discussion**

Effects of ethnic minority share on test scores have been examined in many studies from several countries. In this study we conducted a meta-analysis in order to bring structure into the wide variety of results that have been found so far on this compositional effect. We followed the convention from other studies that compositional effects are treated as the same in nature, but to do justice to existing differences, we also introduced a classification into ethnic groups, both in the type of ethnic group that causes the effect and in the type of group that is affected by the minority share. Distinguishing between ethnic groups that cause the effect seems necessary because important differences exist between ethnic groups in their socioeconomic and historical backgrounds. As a result, increasing the share of students from one minority group, may have stronger effects than increasing the share of students from another minority group. Differentiating between ethnic groups among those affected by the minority share, follows ideas on differential effects that several authors have addressed recently. They argued that school effects often vary between students with different backgrounds. For example, students from poorer socioeconomic background seem more sensitive to school factors in their learning (Muijs et al., 2005; Palardy, 2008). Similar variations in the effect of composition

may be expected between students with different ethnic backgrounds. In a meta-analysis, we can only distinguish between relatively broad ethnic groups, and there may still be variance within these groups that we, out of necessity, ignore. But using different specifications of ordinary meta-regressions and fixed effects meta-regressions, we showed the robustness of our results. This gives confidence in the generalizability of these results and strengthens the belief that there are no large, fundamental differences within the ethnic groups we distinguish. Nevertheless, sub classifications and exceptions may in some cases be very important and should be studied in subsequent work.

The meta-analysis showed that compositional effects on test scores found in the existing literature, are generally not very large, but that there is some important variation. Effects related to the share of African Americans (and effects related to the share of indigenous students) seem considerably stronger than effects related to the share of immigrants. Why exactly this is the case, does not become immediately clear from the reviewed literature and asks for further research. Most likely, it has to do with the situation that African Americans are in, which in a number of aspects is arguably worse than the situation other ethnic groups are in. They face larger achievement gaps than most groups of immigrants and stand out through a long history of overt and legal discrimination. The fact that African American students have such a long history of social, economic and educational deprivation may affect expectations and strengthen stereotypes held by teachers of these students. The causal channel that is largely specific for immigrants, language problems, does not seem to weigh up to this. (Besides, we found no evidence that the effect of immigrant share was stronger on language performance than on mathematics performance.)

The results showed that increasing the ethnic minority share seems to have a stronger effect on students from that same ethnic minority group, than on students belonging to the ethnic majority or on other minority groups. (For immigrant share, the effects on majority's and other minority's test scores even seem close to zero.) Perhaps the effect on the "own" ethnic group is strongest, because students within ethnically mixed schools interact more with students from their own ethnicity than with students from other ethnicities (Echenique & Fryer, 2007). Also, they may be more vulnerable to the way that a teacher's teaching style and expectations change as the minority share increases. Ethnic minority children may also be more vulnerable to compositional effects in general, because of a lower average level of social capital, meaning that they are more dependent for their learning process on the school context.

One issue that we did not address in our meta-analysis is the issue of ethnic composition as a non-linear effect. It is possible that compositional effects get stronger (or weaker) as the ethnic minority share goes up. Consequently, some researchers entered quadratic terms for composition in their regressions, and others compared e.g. schools with less than 25% with schools with 25-50% minority students. Quadratic effects cannot be quantitatively compared to linear effects: they would require a separate meta-analysis. But the number of available studies that satisfies our other criteria for inclusion is too low for this. Results from analyses that compared schools with different shares of minorities cannot easily be compared with each other (e.g. a study using 0-25%, 25-50%, etc. cut-off cannot easily be compared to one using 0-33%, etc.), or with the studies included in the present meta-analysis and therefore had to be excluded. Hence, this meta-analysis does not enable conclusions about the non-linearities that potentially exist in compositional effects.

It is often argued that with regard to compositional effects, it is not ethnicity that matters, but the highly-correlated socio-economic status. We did not find conclusive evidence for this: correcting for average SES does not take away the ethnicity compositional effect. However, this may have to do with unreliable measurement of SES in several of the included studies (see also Van Ewijk & Sleegers, 2007). We also did not find proof for the assumption that compositional effects of ethnicity go via average ability level and neither for the expected differences in effects between composition measured at class- and at cohort- / school-level.

Another often heard argument is that statistically established compositional effects are only artifacts of the method used. The extent to which the studies included in our meta-analysis are at risk of giving biased estimates differs. Studies using simple OLS without a carefully chosen set of covariates will be more at risk than studies using IV or a carefully chosen set of covariates, possibly including some fixed effects. There are a few parameters in our meta-analytic models that touch on this discussion. Not correcting for individual prior achievement when estimating a compositional effect, we argued, would be a major source of bias. However, although point estimates in our models suggested that this bias may be a major concern, these estimates were not significant. Studies defining the effect as peer effect generally had a stronger focus on avoiding bias. They found a bit smaller effects, but this difference was far from significant. Despite this, there are strong arguments about the risks of bias and researchers should always take these into account and carefully choose their models.

For several of the aforementioned parameters, we found large but not significant effects compared to the overall effect. Given the total student N over all

studies which is enormous, we would have expected either small and insignificant, or large and significant effects. That this is not the case and, furthermore, that the variance explained in our models is not very large, has three notable causes. First, the compositional effects were small. So, the parameters are not large in an absolute sense, but large relative to the small primary effects. Second, compositional effects varied between combinations of effects of and on different ethnic groups. For example correcting for average SES or not correcting for pre scores will probably have a larger impact when the compositional effect is large, than when it is around zero. Hence, our assumption that the parameters to several covariates are equal for the different ethnic groups the effect can be of or on, may have been too strong. We could not solve this by adding a set of interactions, since the variance between the included estimates did not suffice to do so. Third, the studies included in our analyses all used multiple regression. This is usually not the case in meta-analyses. In order to capture all differences between the studies, we would have had to include in our meta-regressions a covariate for each covariate used in the studies and for the way in which it was measured. This is not possible. By using meta-regressions instead of univariate meta-ANOVA's, we were able to control for heterogeneity between studies to a substantial extent by adding covariates. But remaining heterogeneity (which is dealt with by adding a random effects variance term) does increase standard errors.

Do our results imply that compositional effects are of no importance and that mixing schools is not a valid policy objective? It may not be that simple. Rumberger & Palardy (2005) have shown that most Afro Americans attend schools where ethnic minorities constitute the majority of the population. Taking Afro American students from a school with, say, a 70% Afro American share to a school with a 20% share, may increase their test score by about 0.13 of a standard deviation; and (as our fixed

effects regressions suggest) even more for older students. Moreover, if this effect works throughout a child's entire school career, it can add up considerably. In similar situations, test scores of immigrants would increase by a lower amount, but the increase would still be of practical importance. Importantly, since the scores of other groups are not affected so much, mixing may lead to a net gain: the average test achievement in a country or area would increase. Even so, the small negative effects found on other ethnic groups, may be held by some as an argument against mixing, as may the argument that effects may not be large enough to warrant difficult and often expensive attempts at mixing. Of course, there may also be entirely different reasons why a society would not want its children to experience schools that are largely segregated by ethnicity. This discussion is a political one that goes beyond the scope of this study. Still research might help politicians to design evidence-based strategies to combat school segregation effectively. Therefore, it is important to carry out meta-analyses on such effects. It makes it possible to bring structure to a diverse field and to see why different studies report varying results on a similar phenomenon. Also, our technique of study fixed effects meta-analysis (which we believe is a valuable addition to current meta-analytic techniques), made it possible to better examine the results reported in a diverse field, using information from differences between multiple estimates given by the same study. We hope that future studies will follow our example in order to create an empirical knowledge base that can help politicians to design evidence-based policies to combat school segregation.

## Appendix 1

### **Adjusting the Weighting Procedure for Multiple Estimates From the Same Study**

When a study contributes several estimates to a meta-analysis, an often argued-for approach to deal with their mutual dependence, is to include only one of the estimates, or to take an average over the estimates (Lipsey & Wilson, 2001). Using this approach here would lead to a loss of valuable information, because estimates from the same study usually differ on some of the covariates. Another approach, proposed by Hox (2002), is to adjust for clustering of data using multilevel meta-analytic models. This is a good approach when a study reports estimates on the same (true) parameter using different subsamples. Here, however, it is not conservative enough, since the estimates often come from exactly the same data on the same (sub)sample; the difference lying in a few covariates. If multiple estimates are made on the same data, then these data determine all estimates that can be made on it at the same time – values on both the predictor of interest and on the dependent variable are the same for each estimate. Hence, we call these estimates “codetermined”. Codetermined estimates do not necessarily come from the same study: if one and the same database of test scores is used in several studies (either by the same or by different authors), then estimates from different studies can also be codetermined. (Note that multilevel meta-analytic models also do not deal with this dependence between estimates from different studies.) Conversely, not all estimates from the same study need to be codetermined: for example an estimate using language test scores comes from a different sub database than an estimate on math scores. If and only if both data on the independent and the dependent variable of interest are the same, we regard estimates as codetermined.

To deal with such estimates, we start from the assumption that we can get no more accurate information from a set of codetermined estimates than the information from the most accurate of the estimates, i.e. the one with the smallest estimation variance,  $v_{ij}^{\text{smallest}}$ . The sum of the inverse estimation variances of all codetermined estimates should not be higher (nor lower) than exactly this. To this end, we divide the smallest estimation variance (“the accuracy”) over all codetermined estimates, proportionally according to the accuracy of each, leading to an adjusted sampling variance for each estimate of:

$$(A1) \quad v_{ij}^* = v_{ij} * v_{ij}^{\text{smallest}} * \sum_{\substack{\text{all code-} \\ \text{terminated} \\ \text{estimates}}} \frac{1}{v_{ij}}$$

This adjusted sampling variance is used in the weighting procedure for the meta-regressions. Note that, following the criteria described above, the majority of the estimates (101) are not codetermined; there are 21 sets of two estimates that are codetermined, four sets of four, two sets of five and one set of eight estimates. Although we believe that adjusting the weighting for codeterminedness is important, the regression parameters and standard errors hardly change if we leave out this correction. An alternative strategy would have been to randomly include only one estimate out of each set of codetermined estimates. This leads to a loss in power and to increased instability in some coefficients. In estimates we made using three different random picks, the main coefficients (effect on / of) proved quite stable (although standard errors increased) while some of the other coefficients started varying in magnitude due to the loss in power and information, although significance levels were unaffected.

## Appendix 2

### **Sensitivity of Results to Removal of Studies Contributing Large Numbers of Estimates**

A potential concern in our analyses may be that a few studies contribute a large number of effect estimates. As discussed in the main text, our results are not expected to depend mainly on these studies. We check this by re-estimating our model after deleting the two studies that contributed the largest number of estimates: Hoxby (2000) and Angrist & Lang (2004). This leads to a loss of power and information, but our regression parameters should not change so much now. After removing the two studies, the number of included effect estimates decreases from 177 to 73. Table A1 shows the results next to the original ones from Table 2. Our results prove reasonably robust against this deletion: most coefficients fall easily within each other's confidence intervals. The effect of age becomes significant and "Effect is *on* students from the ethnic majority" becomes insignificant. Both changes, however, are in line with our fixed effects models. The effect *of* indigenous share (of which Hoxby gave a number of estimates) becomes negative, but stays insignificant. Effect *of* immigrant share and the interaction with language test also show changes. But given that we have deleted quite a bit of valuable information, the number and pattern of noteworthy changes seems acceptable. Our results, thus, show not to be predominantly determined by the two studies contributing the largest number of effect estimates.

-----  
Insert Table A1 about here  
-----

## References

- Angrist, J. D., & Lang, K. (2004). Does School Integration Generate Peer Effects? Evidence from Boston's Metco Program. *American Economic Review*, 94(5), 1613-1634.
- Bankston, C., III, & Caldas, S. J. (1996). Majority African American Schools and Social Injustice: The Influence of De Facto Segregation on Academic Achievement. *Social Forces*, 75(2), 535-555.
- Bankston, C., III, & Caldas, S. J. (1998). Family Structure, Schoolmates, and Racial Inequalities in School Achievement. *Journal of Marriage and the Family*, 60(3), 715-723.
- Burgess, S., & Wilson, D. (2005). Ethnic segregation in England's schools. *Transactions of the Institute of British Geographers*, 30(1), 20-36.
- Caldas, S. J., & Bankston, C., III. (1998). The Inequality of Separation: Racial Composition of Schools and Academic Achievement. *Educational Administration Quarterly*, 34(4), 533-557.
- Caldas, S. J., & Bankston, C., III. (1997). Effect of School Population Socioeconomic Status on Individual Academic Achievement. *Journal of Educational Research*, 90(5), 269-277.
- Canessa, A. (2004). Reproducing Racism: Schooling and Race in Highland Bolivia. *Race Ethnicity and Education*, 7(2), 185-204.
- Card, D., & Rothstein, J. (2006). *Racial Segregation and the Black-White Test Score Gap*. NBER working papers, No. 12078, National Bureau of Economic Research.
- Clotfelter, C., Ladd, H., & Vigdor, J. (2006a). Teacher-Student Matching and the Assessment of Teacher Effectiveness. *Journal of Human Resources*, 41(4), 778-820.
- Clotfelter, C., Ladd, H., & Vigdor, J. (2006b). *The Academic Achievement Gap in Grades 3 to 8*. NBER working papers, No. 12207. National Bureau of Economic Research.
- Cook, T. (1984). *School Desegregation and Black Achievement*. Washington, DC: US Department of Education.
- Cooper, H., & Hedges, L. V. (1994). *The handbook of research synthesis*. New York: Russell Sage Foundation.
- Crain, R.L. & Mahard, R.E. (1983). The Effect of Research Methodology on Desegregation-Achievement Studies: A Meta-Analysis. *The American Journal of Sociology*, 88(5), 839-854.
- De Fraine, B., Van Damme, J., Van Landeghem, G., Opdenakker, M. C., & Onghena, P. (2003). The effect of schools and classes on language achievement. *British educational research journal*, 29(6), 841-860.
- Demack, S., Drew, D., & Grimsley, M. (2000). Minding the Gap: Ethnic, Gender and Social Class Differences in Attainment at 16, 1988-95. *Race Ethnicity and Education*, 3(2), 117-143.
- Driessen, G. (2002). School Composition and Achievement in Primary Education: A Large-Scale Multilevel Approach. *Studies in Educational Evaluation*, 28(4), 347-368.
- Echenique, F., & Fryer, R. G., Jr. (2007). A Measure of Segregation Based on Social Interactions. *The Quarterly Journal of Economics*, 122(2), 441-485.
- Evans, W. N., Oates, W. E., & Schwab, R. M. (1992). Measuring Peer Group Effects: A Study of Teenage Behavior. *Journal of Political Economy*, 100(5).

- Gijsberts, M. (2003). Minderheden in het Basisonderwijs. [Minorities in Primary Education.] In J. Dagevos, M. Gijsberts & C. van Praag (Eds.), *Rapportage Minderheden 2003. Onderwijs, Arbeid en Sociaal-culturele Integratie. [Report Minorities 2003. Education, Labor and Socio-cultural Integration.]* Den Haag, The Netherlands: Sociaal en Cultureel Planbureau.
- Goldhaber, D. D., & Brewer, D. J. (1997). Why Don't Schools and Teachers Seem to Matter? Assessing the Impact of Unobservables on Educational Productivity. *Journal of Human Resources*, 32(3), 505-523.
- Gorard, S., & Smith, E. (2004). An International Comparison of Equity in Education Systems. *Comparative Education*, 40(1), 15-28.
- Gould, E. D., Lavy, V., & Paserman, M. D. (2004). *Does Immigration Affect the Long-Term Educational Outcomes of Natives? Quasi-Experimental Evidence.* NBER working papers. No. 10844. National Bureau of Economic Research.
- Grogger, J. (1996). Does School Quality Explain the Recent Black-White Wage Trend? *Journal of labor economics*, 14(2), 231-253.
- Hanushek, E.A., Kain, J.F., Markman, J.M. & Rivkin, S.G. (2001). *Does peer ability affect student achievement?* NBER working papers. No. 8502. National Bureau of Economic Research.
- Hanushek, E. A., Kain, J. F., & Rivkin, S. G. (2002). *New Evidence about Brown v. Board of Education: The Complex Effects of School Racial Composition on Achievement.* NBER working papers. No. 8741. National Bureau of Economic Research.
- Harker, R., & Nash, R. (1996). Academic Outcomes and School Effectiveness: Type "A" and Type "B" Effects. *New Zealand Journal of Educational Studies*, 31(2), 143-170.
- Harker, R., & Tymms, P. (2004). The Effects of Student Composition on School Outcomes. *School Effectiveness and School Improvement*, 15(2), 177-200.
- Hauser, R. M. (1970). Context and Consex: A Cautionary Tale. *American Journal of Sociology*, 75, 645-664.
- Hedges, L. V., & Vevea, J. L. (1998). Fixed- and Random-Effects Models in Meta-Analysis. *Psychological methods*, 3(4), 486-504.
- Ho Sui Chu, E., & Willms, J. D. (1996). Effects of Parental Involvement on Eighth-Grade Achievement. *Sociology of Education*, 69(2), 126-141.
- Hox, J. (2002). *Multilevel analysis techniques and applications.* Mahwah, NJ: Lawrence Erlbaum Associates.
- Hoxby, C. M. (2000). *Peer effects in the classroom learning from gender and race variation.* NBER working papers. No. 7867. National Bureau of Economic Research.
- Hutchison, D. (2003). The Effect of Group-level Influences on Pupils' Progress in Reading. *British Educational Research Journal*, 29(1), 25-40.
- Jussim, L. & Harber, K.D. (2005), Teacher Expectations and Self-Fulfilling Prophecies: Knowns Unknowns, Resolved and Unresolved Controversies, *Personality and Social Psychology Review*, 9(2), 131-155.
- Katz, L. F., Kling, J. R., and Liebman, J.B. (2001). Moving to Opportunity in Boston: Early results of a randomized mobility experiment. *Quarterly Journal of Economics* 116(2), 607-654.
- Lauder, H., & Hughes, D. (1999). *Trading in Futures: Why Markets in Education Don't Work.* Philadelphia: Open University Press.
- Lee, V. E., & Bryk, A. S. (1989). A Multilevel Model of the Social Distribution of High School Achievement. *Sociology of Education*, 62(3), p172-192.

- Link, C. R., & Mulligan, J. G. (1991). Classmates' Effects on Black Student Achievement in Public School Classrooms. *Economics of Education Review*, 10(4), 297-310.
- Lipsey, M. W., & Wilson, D. B. (2001). *Practical meta-analysis*. Thousand Oaks, CA.: Sage.
- McEwan, P. J. (2003). Peer Effects on Student Achievement: Evidence from Chile. *Economics of Education Review*, 22(2), 131-141.
- McEwan, P. J. (2004). The Indigenous Test Score Gap in Bolivia and Chile. *Economic Development and Cultural Change*, 53(1), 157-190.
- Meredith, Custodial Parent and Next Friend of McDonald v. Jefferson County Bd. of Ed et al., No. 05-915, (2007).
- Muijs, D., Campbell, R., Kyriakides, L., & Robinson, W. (2005). Making the Case for Differential Teacher Effectiveness: An Overview of Research in Four Key Areas. *School Effectiveness and School Improvement*, 16(1), 51-70.
- Musterd, S. (2005). Social and Ethnic Segregation in Europe: Levels, Causes, and Effects. *Journal of Urban Affairs*, 27(3), 331-348.
- Nash, R. (2003). Is the School Composition Effect Real? A Discussion with Evidence from the UK PISA Data. *School Effectiveness and School Improvement*, 14(4), 441-457.
- OECD. (2001). *Knowledge and Skills for Life: First Results from the OECD Programme for International Student Assessment (PISA), 2000*. (No. 92-64-19671-4). Paris: OECD Publications.
- OECD. (2003). *Literacy skills for the world tomorrow: further results from PISA 2000* (No. 9264102868). Paris: OECD.
- Opdenakker, M. C., Van Damme, J., De Fraine, B., Van Landeghem, G., & Onghena, P. (2002). The Effect of Schools and Classes on Mathematics Achievement. *School Effectiveness and School Improvement*, 13(4), 399-427.
- Overton, R. C. (1998). A Comparison of Fixed-Effects and Mixed (Random-Effects) Models for Meta-Analysis Tests of Moderator Variable Effects. *Psychological methods*, 3(3), 354-379.
- Palardy, G. J. (2008). Differential school effects among low, middle, and high social class composition schools: A multilevel, multiple group latent growth curve analysis. *School Effectiveness and School Improvement*, 19(1), 21-49.
- Parents Involved in Community Schools v. Seattle School District No. 1 et al., No. 05-908 (2007).
- Peetsma, T., Van der Veen, I., Koopman, P., & Van Schooten, E. (2005). *Class composition influences on pupils' cognitive development*. Amsterdam, The Netherlands: University of Amsterdam, SCO-Kohnstamm Institute, Working Paper.
- Raudenbush, S.W. (1994). Random effect models. In: H. Cooper & L.V. Hedges (Eds.), *The handbook of research synthesis* (pp. 301-321), Russell Sage Foundation, New York.
- Riach, P.A. & Rich, J. (2002). Field experiments of discrimination in the market place. *The Economic Journal*, 112, F480-F518.
- Rivkin, S. G. (2000). School Desegregation, Academic Attainment, and Earnings. *Journal of Human Resources*, 35(2), 333-346.
- Rivkin, S. G. (2001). Tiebout Sorting, Aggregation and the Estimation of Peer Group Effects. *Economics of Education Review*, 20(3), 201-209.

- Rosenthal, R. & Jacobson, L. (1968). *Pygmalion in the classroom: Teacher expectations and student intellectual development*. New York: Holt, Rinehart & Winston.
- Rumberger, R. W., & Palardy, G. J. (2004). Multilevel Models for School Effectiveness Research. In D. Kaplan (Ed.), *Handbook of Quantitative Methodology for the Social Sciences* (pp. 235-258). Thousand Oakes, CA: Sage.
- 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.
- Sacerdote, B. (2001). Peer Effects with Random Assignment: Results for Dartmouth Roommates. *Quarterly Journal of Economics*, 116(2), 681-704.
- Scafidi, B., Sjoquist, D. L., & Stinebrickner, T. R. (2007). Race, poverty, and teacher mobility. *Economics of education review*, 26(2), 145-159.
- Schindler-Rangvid, B. (2007). Living and Learning Separately? Ethnic Segregation of School Children in Copenhagen. *Urban studies*, 44(7), 1329-1354.
- Stevens, P. A. J. (2007). Researching Race/Ethnicity and Educational Inequality in English Secondary Schools: A Critical Review of the Research Literature Between 1980 and 2005. *Review of Educational Research*, 77(2), 147-185.
- Strand, S. (1997). Pupil Progress during Key Stage 1: A value added analysis of school effects. *British educational research journal*, 23(4), 471-488.
- Thernstrom, A., & Thernstrom, S. (2003). *No Excuses: Closing the Racial Gap in Learning*. New York: Simon & Schuster.
- Thompson, S. G., & Higgins, J. P. T. (2002). How should meta-regression analyses be undertaken and interpreted? *Statistics in medicine*, 21(11), 1559-1574.
- Thrupp, M. (1995). The School Mix Effect: The history of an enduring problem in educational research, policy and practice. *British Journal of Sociology of Education*, 16(2), 183-204.
- Van Ewijk, R. J. G. & Slegers, P. J. C. (2007). *The effect of peer SES on achievement: a meta-analysis*. Paper presented at the 2007 annual meeting of the American Educational Research Association, Chicago, IL.
- Vigdor, J., & Nechyba, T. (2004). *Peer Effects in North Carolina Public Schools*. Durham, USA: Duke University, Working Paper.
- Willms, J. D. (1986). Social Class Segregation and its Relationship to Pupils' Examination Results in Scotland. *American sociological review*, 51(2), 224-241.
- Wortman, P.M. & Bryant, F. B. (1985). School Desegregation and Black Achievement. *Sociological Methods & Research*, 13(3), 289-324.
- Zimmerman, D. J. (2003). Peer Effects in Academic Outcomes: Evidence from a Natural Experiment. *The review of economics and statistics*, 85(1), 9-23.

Table 1. Summary of studies included in the meta-analysis.

Author(s) (publication year)	Estimates in basic meta-regressions	Estimates in fixed effects models	Country	Effect of → on	Effect size: minimum / maximum for combination
Angrist & Lang (2004)	32	32	USA	AA → same AA → all	-6.24 / 1.13 (OLS); -11.15 / -.42 (IV) -.41 / .15 (OLS); -1.12 / -.66 (IV)
Bankston & Caldas (1996)	12	12	USA	AA → same AA → all AA → maj	-.68 / -.37 -.49 / -.30 -.32 / -.13
Bankston & Caldas (1998)	2	2	USA	AA → all	-.28 / -.26
Caldas & Bankston (1997)	1	-	USA	AA → all	-.48
Caldas & Bankston (1998)	3	3	USA	AA → same AA → all AA → maj	-.25 -.24 -.28
Harker & Nash (1996)	9	9	New Zealand	imm → all ind → all	-1.86 / .76 -.61 / .17
Hoxby (2000)	72	72	USA	AA → same AA → imm AA → maj imm → maj imm → same imm → other	-1.07 / .10 -.86 / -.18 -.30 / .08 -.19 / .03 -.63 / .02 -.18 / .17

				ind → other	-3.56 / 1.73
				ind → maj	-1.20 / 1.53
McEwan (2003)	6	6	Chile	ind → all	-.30 / -.30 (school fixed effects) -2.77 / 2.70 (siblings & twins fixed effects)
McEwan (2004)	8	8	Chile & Bolivia	ind → all	-.33 / .11
Peetsma et al. (2005)	22	22	Netherlands	imm → maj	-.21 / -.16
				imm → all	-.15 / .03
				imm → other	-.20 / .36
Rivkin (2000)	1	-	USA	AA → same	.20
Strand (1997)	1	-	UK	imm → all	-1.03
Vigdor & Nechyba (2004)	8	8	USA	AA → same	-.10 / -.05
				AA → maj	-.13 / -.05
				imm → maj	-.12 / -.12
				imm → same	.04 / .19

---

Note: AA = African Americans; imm = immigrants; ind = indigenous; all = entire student population; same = effect is on students from same ethnic minority; other = effect is on students from another ethnic minority.

Table 2. Parameter estimates and (standard errors) for the first set of basic meta-regression models: general differential effects.

	(1)	(2)	(3)	(4)
Constant (= effect of African American share on students from the same ethnic minority)	-0.178 ** (0.020)	-0.364 ** (0.048)	-0.363 ** (0.057)	-0.271 ** (0.089)
Effect is of immigrant share		0.161 ** (0.044)	0.179 ** (0.063)	0.133 * (0.061)
Effect is of indigenous share		0.056 (0.069)	0.047 (0.070)	0.014 (0.089)
Effect is on entire student population		0.125 (0.068)	0.132 (0.069)	0.121 (0.067)
Effect is on students from the ethnic majority		0.128 * (0.055)	0.124 * (0.055)	0.117 * (0.049)
Effect is on students from another ethnic minority		0.180 * (0.070)	0.172 * (0.070)	0.153 * (0.067)
Age (centered at 12)			-0.006 (0.010)	-0.005 (0.015)
Test = language			0.004 (0.058)	0.011 (0.052)
Interaction: test = language x effect of immigrants			-0.058 (0.083)	-0.062 (0.074)
Studied effect = peer effect (ref. cat: compositional effect)				0.059 (0.112)
No individual prior achievement covariate included				-0.176 (0.169)
Composition measured at cohort or school level (ref. cat.: at class level)				0.015 (0.088)
Model included covariate for average SES				0.063 (0.060)
Model included covariate for average prior achievement / ability				-0.005 (0.140)
R <sup>2</sup>	n.a.	0.109	0.113	0.163
Systematic variance component ( $\sigma_{\theta}^2$ )	0.0221 ** (0.0057)	0.0192 ** (0.0052)	0.0182 ** (0.0050)	0.0119 ** (0.0038)

Note: \* = significant at .05 level; \*\* = significant at .01 level

Table 3. Parameter estimates and (standard errors) for the second basic meta-regression model: differential effects specified.

Constant (= effect of African American share on African American students)	-0.265 ** (0.088)
Effect is of African American share on entire student population	0.157 (0.088)
Effect is of African American share on students from another ethnic minority (viz. Hispanics)	-0.053 (0.091)
Effect is of African American share on students from the ethnic majority	0.130 * (0.061)
Effect is of immigrant share on students from the ethnic majority	0.181 * (0.076)
Effect is of immigrant share on entire student population	0.154 (0.109)
Effect is of immigrant share on students from the same ethnic minority	0.070 (0.087)
Effect is of immigrant share on students from another ethnic minority	0.398 ** (0.094)
Effect is of indigenous share on entire student population	0.127 (0.085)
Age (centered at 12)	-0.011 (0.014)
Test = language	0.011 (0.047)
Interaction: test = language x effect of immigrants	-0.060 (0.069)
Studied effect = peer effect (ref. cat: compositional effect)	0.067 (0.115)
No individual prior achievement covariate included	-0.196 (0.174)
Composition measured at cohort or school level (ref. cat.: at class level)	0.054 (0.088)
Model included covariate for average SES	0.044 (0.059)
Model included covariate for average prior achievement / ability	-0.007 (0.138)
R <sup>2</sup>	0.255
Systematic variance component ( $\sigma_{\theta}^2$ )	0.0082 ** (0.0031)

Note: \* = significant at .05 level; \*\* = significant at .01 level

Table 4. Added-up estimates from meta-regression I and II; number of contributed estimates per combination from meta-regression II.

Effect of increasing the share of	Effect on test scores of	Estimate from meta-regression I	Estimate from meta-regression II	Number of studies and (effect estimates) contributed in meta-regression II <sup>14</sup>
African Americans	students from the same ethnic minority	-0.271	-0.265	5 (30)
African Americans	students from another ethnic minority	-0.118	-0.318	1 (8)
African Americans	entire student population	-0.150	-0.108	5 (22)
African Americans	students from the ethnic majority	-0.155	-0.135	4 (15)
Immigrants	students from the same ethnic minority	-0.139	-0.195	3 (14)
Immigrants	students from another ethnic minority	0.014	0.133	4 (23)
Immigrants	entire student population	-0.018	-0.111	4 (11)
Immigrants	students from the ethnic majority	-0.022	-0.084	3 (13)
Indigenous	students from the same ethnic minority	-0.257		
Indigenous	students from another ethnic minority	-0.104		
Indigenous	entire student population	-0.136	-0.138	4 (41)
Indigenous	students from the ethnic majority	-0.140		

Note: estimates from meta-regression I are derived from the last column of Table 2. The estimates from meta-regression II are those from Table 3. The estimates refer to effects on math scores of 12-year old pupils; all covariates for study and model characteristics not relating to the ethnic group the effect is of or on (see Table 2 and 3) are set to zero.

Table 5. Parameter estimates and (standard errors) for the study fixed effects meta-regression model, following the first basic meta-regression model (general differential effects).

---

Effect is <i>of</i> immigrant share	0.146 ** (0.044)
Effect is <i>of</i> indigenous share	-0.084 (0.115)
Effect is <i>on</i> students from the ethnic majority	0.049 (0.028)
Effect is <i>on</i> students from another ethnic minority	0.089 (0.049)
Test = language	0.015 (0.029)
Age	-0.044 ** (0.011)
Interaction: test = language x effect of immigrants	-0.074 (0.047)
Model included covariate for average SES	-0.003 (0.056)

---

Note: \* = significant at .05 level; \*\* = significant at .01 level. Predictors that were included in the ordinary meta-regression, but are not shown above, could not be included in the study fixed effects meta-regression, because of lacking within-study variation.

Table 6. Parameter estimates and (standard errors) for the fixed effects meta-regression model, following the second basic meta-regression model (differential effects specified).

---

Effect is <i>of</i> African American share <i>on</i> entire student population	0.060 (0.071)
Effect is <i>of</i> African American share <i>on</i> students from another ethnic minority (viz. Hispanics)	-0.085 (0.073)
Effect is <i>of</i> African American share <i>on</i> students from the ethnic majority	0.058 (0.036)
Effect is <i>of</i> immigrant share <i>on</i> students from the ethnic majority	0.164** (0.053)
Effect is <i>of</i> immigrant share <i>on</i> entire student population	0.204* (0.082)
Effect is <i>of</i> immigrant share <i>on</i> students from the same ethnic minority	0.050 (0.067)
Effect is <i>of</i> immigrant share <i>on</i> students from another ethnic minority	0.386** (0.075)
Effect is <i>of</i> indigenous share <i>on</i> entire student population	0.248 (0.390)
Age	-0.044** (0.011)
Test = language	0.017 (0.029)
Interaction: test = language x effect of immigrants	-0.075 (0.047)
Model included covariate for average SES	0.009 (0.062)

---

Note: \* = significant at .05 level; \*\* = significant at .01 level. Predictors that were included in the ordinary meta-regression, but are not shown above, could not be included in the study fixed effects meta-regression, because of lacking within-study variation.

Table A1. Comparison of models with and without the two studies contributing the largest numbers of effect estimates.

	Final model (Table 2)	Without Hoxby (2000) and Angrist & Lang (2004)
Constant (= effect of African American share on students from the same ethnic minority)	-0.271 ** (0.089)	-0.189 * (0.091)
Effect is of immigrant share	0.133 * (0.061)	0.044 (0.085)
Effect is of indigenous share	0.014 (0.089)	-0.167 (0.140)
Effect is on entire student population	0.121 (0.067)	0.037 (0.053)
Effect is on students from the ethnic majority	0.114 * (0.049)	-0.014 (0.039)
Effect is on students from another ethnic minority	0.153 * (0.067)	0.242 ** (0.077)
Age (centered at 12)	-0.005 (0.015)	-0.046 ** (0.013)
Test = language	0.011 (0.052)	0.021 (0.034)
Interaction: test = language x effect of immigrants	-0.062 (0.074)	-0.163 ** (0.062)
Studied effect = peer effect (ref. cat: compositional effect)	0.059 (0.112)	0.083 (0.149)
No individual prior achievement covariate included	-0.176 (0.169)	-0.013 (0.170)
Composition measured at cohort or school level (ref. cat.: at class level)	0.015 (0.088)	0.088 (0.097)
Model included covariate for average SES	0.063 (0.060)	-0.001 (0.051)
Model included covariate for average prior achievement / ability	-0.005 (0.140)	-0.019 (0.161)
$R^2$	0.163	0.650
Systematic variance component ( $\sigma_\theta^2$ )	0.0119 ** (0.0038)	0.0000 (0.0010)

Note: \* = significant at .05 level; \*\* = significant at .01 level.

## Footnotes

- <sup>1</sup>In many countries, there are also groups of immigrants that are doing no worse on important indicators, such as inter-European Union immigrants in the EU and East-Asians in the USA. Generally, these groups are not an important concern to public and politicians, nor are compositional effects related to them a source of scientific debate. This study will only focus on ethnic minorities in various countries that are in disadvantaged situations, relative to the country's ethnic majority.
- <sup>2</sup>Some immigrants have the same native language as the main language in their host country. In the studies included in our meta-analysis, however, this only applies to a minor part of the immigrants in The Netherlands in Peetsma et al. (2005).
- <sup>3</sup>Earlier meta-analyses in this domain have been carried out in the 1980s (Cook, 1984; Crain & Mahard, 1983; Wortman & Bryant, 1985). Since they only focused on effects of lifting de jure segregation in the United States and because of the generally poorer quality of the then-included studies, their general conclusion that desegregation had a small positive effect on African Americans' reading levels, especially if desegregation took place in the earliest grades and that desegregation did not have a negative effect (but also little positive effect) on African Americans' math achievement, cannot be generalized to effects in the current situation. Also, our scope is broader, as we include countries other than the USA and do not only look at effects on students from the ethnic minority.
- <sup>4</sup>In a few cases, no standard error was reported, but it was only mentioned that the effect was significant at e.g.  $p = .05$ . We then calculated the standard error (conservatively) assuming a  $p$  of  $.05$  and analogously for other significance levels (cf. Cooper & Hedges, 1994). Some studies reported standard errors that were incorrect, because OLS was used without taking into account the clustered nature of the data. These standard errors were adjusted based on the distribution of variances over class/cohort and school (which, if not available, was estimated from studies using similar datasets) and group sizes. In a few studies, some compositional effects had to be calculated as the sum of a main effect (effect of minority share) and an interaction effect (minority share \* a dummy for own ethnicity). In these cases, the true standard error is only calculable if the original data's variance-covariance data is known, which is generally not the case. Instead, the standard error to the interaction parameter was taken. If  $X_1$  gives the proportion ethnic minorities,  $X_2$  is a dummy variable which indicates whether a student is from the ethnic minority,  $X_3$  gives the interaction (i.e.  $X_1 * X_2$ ), and  $\beta_1$ ,  $\beta_2$  and  $\beta_3$  are the regression parameters to these variables, it can, although it would be too space-consuming here, easily be proven that  $SE(\beta_1 + \beta_2)$  (the true standard error we would like to know) is always smaller than  $SE(\beta_3)$  (the standard error we take), so that the standard error we take is an overestimation of the true, unknown standard error. Hence, this is a conservative approach.
- <sup>5</sup>Note that "fixed effect" here refers to something that is entirely different from what is generally meant by fixed effects meta-regression (cf. Cooper & Hedges, 1994; Lipsey & Wilson, 2001): such models are similar to our equation (1), but omit error term  $u_{ij}$ . Whenever we mention fixed effects, we do not refer to this latter type of model.
- <sup>6</sup>Since all systematic between-study variation is captured in  $\alpha_j$ , it is not necessary to include a random-effects error term  $u_{ij}$  here.
- <sup>7</sup>Some studies measured composition at the cohort level, but the number of classes per cohort averaged little over one, so that cohort and class essentially were the same. We classified studies that measured composition in units of on average 40 students or less as measuring at class level. Studies measuring composition in larger units, were classified as cohort / school level. (Note that in countries of the OECD, average secondary school class sizes are up to 39 (OECD, 2003).
- <sup>8</sup>Only one study, Harker & Nash (1996) gave a few estimates on science test scores. The sizes of their coefficients roughly lay between those for math and those for language. Since the number of estimates for science was too small to use a separate category for them, we grouped them with the math estimates, in order to be able to compare the effects of language to those on other tests. Science shares with mathematics that it is a technical subject, but its content is much more language-dependent. Alternatively, the science estimates could have been grouped with the language estimates or be omitted from the meta-analyses. We also estimated these alternative models. Neither lead to any meaningful changes in coefficients.

- <sup>9</sup>Especially for the covariates “peer” and inclusion of average SES and ability, one might argue for another “best” category. We follow the interpretation here that the effects studied are compositional effects that do include those parts of the effect going via SES and ability. The insignificance of these parameters shows that another way of dummy-coding would not have influenced our results in an important way.
- <sup>10</sup>In an alternative estimation, we replaced the division between studies on peer effects and studies on compositional / contextual effects by a division into those studies that used a formal strategy to deal with omitted variables (such as instrumental variables and panel data with fixed effects) and those that did not. The parameter to this dummy variable did not become significant. Some other point estimates changed somewhat, but interpretations remained the same and the parameter estimates to the effect of / on parameters were not affected to any important extent.
- <sup>11</sup> Also of influence may be the estimates from Hoxby (2000): in her models, not including a prior achievement covariate will not lead to the same bias as in other studies, because of the very different estimation strategy that she uses.
- <sup>12</sup>Only Hoxby (2000) gives some estimates of indigenous share on test scores of African Americans, Hispanics and Whites (but not on test scores of indigenous pupils themselves). To avoid parameters in this meta-regression to be determined by only one study and to increase the power of the “effect of indigenous on entire population” parameter, we combine Hoxby’s estimates by means of a regular fixed effects meta-regression into an effect of indigenous share on test scores of the entire population.
- <sup>13</sup>We should note here that the African American on other minority effect is entirely derived from Hoxby (2000). Although the estimate in the table suggests that this effect is not weaker, and perhaps even stronger than the African American on African American effect, Hoxby actually reports the latter effect to be clearly stronger. This may seem confusing, but can easily be explained by the fact that our African American on African American parameter is a weighted average over several studies, while our African American on other minority parameter is estimated using only Hoxby’s study. The difference is now in the other direction and not significant.
- <sup>14</sup>Note: a few studies gave estimates that could not be classified exactly into one of the categories, for example an effect of African American share on scores of a combined group of Whites and Hispanics in Vigdor & Nechyba (2004). Such estimates were classified proportionally, in this case as 0.96 times effect of African Americans on Whites and as 0.04 times effect of African Americans on Hispanics. In counting the contributed number of effect estimates shown in the table, the sum of all (complete and proportional) contributions was rounded to the nearest integer. In counting the number of contributing studies, a study was only counted if the sum of proportions for a certain combination (e.g. in the case of Vigdor & Nechyba, two estimates times 0.04 makes 0.08) was at least 0.50, so that the study cumulatively contributed at least half an effect estimate.