Peer effects in European primary schools: Evidence from PIRLS

Andreas Ammermueller
ZEW, Mannheim

Jörn-Steffen Pischke
LSE, London

January 2006

Abstract

We estimate peer effects for fourth graders in six European countries. The identification relies on variation across classes within schools. We argue that classes within primary schools are formed roughly randomly with respect to family background. Similar to previous studies, we find a modest but significant impact of peer background on reading test scores. We find no significant evidence on non-linear peer effects. The results indicate that sorting across schools is strong and cannot be fully controlled for by available class and school characteristics. This leads to an upward bias in standard estimates of peer effects not controlling for school effects.

JEL classification: I21

We thank David Autor and Matt Pritsker for helpful comments. Ammermueller greatfully acknowledges financial support from the Fritz Thyssen Foundation under the project "Bildungschancen zwischen Grundschule und Sekundarstufe."

1 Introduction

Learning in schools takes place in a group setting, and the composition of the group possibly affects individual outcomes. There has been a lot of interest in these types of social interactions in economics recently, and in peer effects in school in particular. We revisit this issue in this paper, drawing on a previously unexploited data set in this context, the Progress in International Reading Literacy Study (PIRLS) for fourth graders. Our analysis covers six European countries, Germany, France, Iceland, the Netherlands, Norway, and Sweden.

Peer effects are of interest for a variety of reasons. From the policy perspective, it is important to know how students should be grouped in schools and class rooms in order to maximize achievement, or to achieve certain equity goals. Tracking policies, the explicit grouping of students into classes by ability is widespread in many countries. Policies increasing school choice and school competition are being discussed or implemented in many countries. These policies will influence student outcomes through peer effects, if there is increased sorting of students as a result of these policies. Understanding peer effects is therefore an important ingredient in evaluating many education policies.

One of the main challenges in the literature on peer effects is the feature that schools and class rooms are not formed randomly. School and class composition typically reflects neighborhood characteristics, and therefore the family background of students. The estimated peer effect may capture unobserved aspects of an individual student's performance if this problem is ignored. We exploit the fact that the PIRLS data sample multiple class rooms within a single school. This allows us to estimate peer effects within schools. Since we study students in primary schools, there is no explicit tracking in any of the countries in our sample. We argue that classes within schools are in fact formed more or less randomly with respect to family background characteristics (other than immigrant status). The variation in our peer variable therefore most likely reflects that there will be small differences in composition when multiple groups are formed out of a small population (in essence the absence of the law of large numbers). Hence, our research design allows for a relatively credible identification of peer effects on student test scores.

The existing literature, of which some recent studies are summarized in Table 1, has used a wide variety of approaches to identify peer effects. The papers closest in spirit to ours are the ones by Gould et al. (2005) for Israel and Hoxby (2000) for the US. These papers similarly rely on differences in the compositions of individual class rooms within a school, which come about by chance. Hanushek et al. (2003) and McEwan (2003) also use within school variation to identify peer effects. However, it is more difficult to believe that differences in class room composition are random in their cases. We will compare our methodology in detail to the existing literature in the next section. A number of recent studies have also used explicit random assignment to classes or schools, or other natural experiments. However, none of these studies is for European countries.

On average across countries, we find that a one standard deviation change in our measure of peer composition leads to a 0.07 standard deviation change in reading test scores, and this estimate is marginally significant. The size of this effect is in the middle of the range of estimates reported in the literature. We find a much larger association between peer characteristics and student outcomes when we do not control for school effects. This suggests that sorting across schools is very important, and cannot be controlled for with standard observable characteristics. Studies which do not address this issue will likely find misleading estimates.

The remainder of the paper is structured as follows. The next section discusses our methodology and compares our approach to the existing literature. Section 3 describes the PIRLS data. In section 4, we investigate the issue of whether classes are formed randomly within schools, and section 5 presents our estimates of peer effects. A final section concludes.

2 Empirical framework and existing literature

A peer effects study typically starts from a specification of an education production function like

$$y_{ics} = \alpha + \beta X_{ics} + \gamma S_{cs} + \delta \overline{X}_{(-i)cs} + \phi \overline{y}_{(-i)cs} + \eta_{cs} + \varepsilon_{ics} \quad (1)$$

where y_{ics} is a student outcome, like a test score, for student i in class room c and school s, X_{ics} are student or family characteristics, like sex, family background, etc., S_{cs} are school or class level characteristics, like class size, teacher experience, characteristics of the municipality, etc., $\overline{X}_{(-i)cs}$ are the average characteristics of the peers of student i, and $\overline{y}_{(-i)cs}$ is the average outcome of the peers. In addition, η_{cs} and ε_{ics} are a class level and an individual level error term.

In the language of Manski (1993), the coefficient δ reflects exogenous or contextual effects, ϕ reflects endogenous effects, and η_{cs} reflects correlated effects. Exogenous effects arise when individuals learn more because the group of peers is more favorable in terms of their predetermined characteristics. Correlated effects arise when the group of peers is subject to a common influence, which is not modeled directly. These effects will give rise to a bias if they are correlated with peer group composition. For example, consider a remedial class room with relatively poorly performing children. This class room may be assigned a particularly able teacher but the exceptional characteristics of this teacher are not observable. Removing the potential bias from contextual effects is one of the main challenges in the peer effects literature, and we will discuss this issue shortly. Endogenous effects arise when individuals learn more because peers are learning more.

Even if the requisite data are available, eq. (1) cannot be estimated directly because of the well known reflection problem (see Manski, 1993). Student i influences his or her peers, and the peers influence student i in turn through the coefficient ϕ, if a symmetric equation to (1) holds for every student in the class. This implies that ε_{ics} will be mechanically correlated with $\overline{y}_{(-i)cs}$, leading to a classic simultaneity problem. It is very difficult to overcome the reflection problem without severe restrictions on the model. The literature has therefore basically resorted to estimating the reduced form equation

$$\begin{aligned} y_{ics} &= \frac{\alpha}{1-\phi} + \beta X_{ics} + \gamma S_{cs} + \frac{\delta + \beta\phi}{1-\phi}\overline{X}_{(-i)cs} + \eta_{cs} + \varepsilon_{ics} \\ &= \alpha' + \beta X_{ics} + \gamma S_{cs} + \lambda \overline{X}_{(-i)cs} + \eta_{cs} + \varepsilon_{ics} \end{aligned} \tag{2}$$

which is the approach we will follow as well. In eq. (2), it is not possible to distinguish endogenous and exogenous effects anymore; either one would give rise to a positive coefficient λ on the peer variable $\overline{X}_{(-i)cs}$. While distinguishing endogenous and exogenous effects may be of interest from a theoretical perspective, for policy purposes it will often not matter which give rise to peer effects. Hence, identifying the reduced form parameter is still of substantial interest.

Now consider the problem arising from the presence of correlated effects in eqs. (1) and (2). If some relevant school or class room characteristics are not controlled, the estimated peer effect λ will be biased. Random assignment of students and teachers to class rooms solves this problem, because random assignment breaks the link between peer characteristics and extraneous effects on the class, like unobserved teacher quality. Boozer and Cacciola (2001) and Graham (2004) exploit the random assignment in the Tennessee STAR experiment on class size. Cullen, Jacob, and Levitt (2003) use lotteries at oversubscribed Chicago public schools. However, their paper does not focus on the issue of peer effects.

True random assignment variation is rare in an education context, and unavailable in many countries. Hence, researchers have to resort to other strategies utilizing the existing data. In this paper, as in a variety of related studies, we use variation within schools in order to identify the peer effect. This means that we estimate

$$y_{ics} = \alpha_s + \beta X_{ics} + \gamma S_{cs} + \lambda \overline{X}_{(-i)cs} + \eta_{cs} + \varepsilon_{ics} \tag{3}$$

instead of eq. (2). α_s is a set of dummy variables for each school. Alternatively, we also estimate an equation with peer variables at the school level. The idea behind this strategy is the observation that different schools draw students from different neighborhoods, and hence family backgrounds. However, in primary school students are not generally grouped into classes on the basis of ability or family background. Although some countries, like Germany, track students into a rigid system of separate schools at the secondary level, there is no system wide tracking at the primary level. In fact, classes in primary schools with multiple class rooms at the same grade level are typically formed more or less on a random basis. In this case $\overline{X}_{(-i)cs}$ will be uncorrelated with the class level shocks η_{cs} conditional on a set of school fixed effects. The bias from correlated effects is thus removed and λ can be estimated consistently. A necessary condition for this to work is, of course, that there is sufficient variance in peer composition of a class room within a school. We will demonstrate that this is true in our data in section 4.

Our identification strategy is most closely related to that of McEwan (2003). He studies peer effects for 8^{th} graders in Chile. However, random assignment to classes within schools is much less likely to happen at the secondary level because schools in many countries, including Chile, track students to at least some degree. If there is tracking on the basis of (unobserved) ability, estimates of λ are still confounded by correlated effects.

Gould et al. (2005), Hanushek et al. (2003), and Hoxby (2000) also use within school variation to identify the peer effects. The Gould et al. and Hoxby studies are most similar in spirit to ours. We use comparisons across class rooms within the same grade for the same cohort of students. Hoxby uses comparisons between classes in the same grade across adjacent cohorts and years. Hence, she identifies peer effects from variation arising from the composition of subsequent cohorts. For example, one cohort may have more girls and the next cohort fewer for purely random reasons. Gould et al. also use data on multiple cohorts in the same grade. They condition on the student composition of the grade across years. Effectively, like Hoxby, they therefore exploit year to year variation in the composition of a cohort of students. However, these studies tend to focus on different peer group measures than ours. Hoxby looks at gender and race composition of the class room and performance by opposite gender and race groups, while Gould et al. look at the share of immigrants.

Hanushek et al. (2003) focus on a peer measure more similar to ours. They also control for school by grade effects like Hoxby and Gould et al. However, they track the same cohort of students over time, rather than different cohorts, and they control for student fixed effects. This means that they effectively only consider changes in the peer group, which come about through student mobility, because classes are typically kept together in primary schools. This means that an individual's peer group only changes if the individual herself or some peers move in or out of the classroom. But student mobility might well be related to unobserved time and grade specific shocks. For example, if a class is about to be assigned a particularly bad teacher, the parents of the best students may move their children to other schools. Peer quality in the class would fall. Student achievement may also subsequently fall in the class room but this might simply be due to the poor teacher, not the poor peers. Hence, controlling for student fixed effects may lead to an upward bias in the estimates. In fact, Hanushek et al.

find an increase in the peer coefficient when they control for individual student effects compared to a similar specification without individual effects.

Table 1 summarizes these and other recent studies on peer effects in schools. These previous literature finds peer effects which range from close to zero (Angrist and Lang, 2004, and Cullen, Jacob and Levitt, 2003) to modestly large (Hoxby, 2000).

3 Data and descriptive statistics

Thirty-five countries participated in the Progress in International Reading Literacy Study (PIRLS). This study was conducted by the International Association for the Evaluation of Educational Achievement (IEA) in 2001 and nine- and ten-year-olds in reading literacy. In the data, extensive information on home and school environment is available through student, parent, teacher and school questionnaires. With 150,000 students tested, PIRLS 2001 is the first in a planned 5-year cycle of international trend studies in reading literacy (Mullis et al., 2003).

The data are collected in a two-stage stratified sampling design. First, participating schools were chosen. Therefore, the schools are the primary sampling units and not the classes or students. Within each school, a sample of classes from the targeted grade was drawn. The targeted grade is the upper of the two grades with the most 9 year-olds at the time of testing. This is always the fourth grade in our sample of countries. Within each class, in principle, all students are sampled. In practice, the number of sampled students can be smaller than the actual class size, because of student non-participation. We use all European countries with a sufficient number of schools with at least two classes. These are France, Germany, Iceland, the Netherlands, Norway and Sweden.

Student performance is measured by test scores in reading literacy, which is the most important basic competency needed to acquire further skills and knowledge and to successfully participate in social life (Mullis et al., 2003). The test scores are plausible values that are drawn from an estimated proficiency distribution. Plausible values are imputed scores based on the students' answers to the test items (cf. Mislevy, 1991). The scores have then been standardized to an international mean of 500 and a standard deviation of 100, which facilitates the comparison across countries.

Table 2 provides information on mean reading scores and sample sizes in PIRLS at the student, class and school level. Students, classes and schools can be directly identified. Missing values of student background, class, and school variables are a serious problem in the data set. For parents' education, 36 percent of all values are missing. Instead of parents' education, we use the number of books at home as our indicator of family background. Among the variables reflecting family background, this is the one with best item response rate. In addition, this is an appealing variable in its own right. It is clearly correlated with parental income, education and origin. Table A3 presents the weighted means of student background variables by categories of the books at home variable. In addition, it reflects whether the parents value literary skills. Parents who own many books most likely will also promote reading among their children. In fact, Woessmann (2004) found the number of books to be the single most important predictor of reading skills among various family background variables in the Third International Math and Science Study (TIMSS) and Ammermueller (2005) in PIRLS and the Programme for International Student Assessment (PISA) data. As an alternative peer measure, we use an index of parents' highest educational attainment and report the results in the Appendix.

Table 2 demonstrates that the sample size, conditioning on non-missing student background and school variables, shrinks to about 40 to 75 percent of the original. Row 4 in the table gives the sizes of the samples we actually use. All figures from row four onward refer to the sample with no missing values. Reading scores in the selected samples are slightly higher than in the overall sample. Some sample schools have only one class. Our within school estimates will only be utilizing the schools with two or more classes. Information on the students, classes, and schools with more than one class can be found in the bottom rows of Table 2. Means and standard deviations of all the variables are reported in Table A1, while Table A2 presents the percentage of missing values for each variable in the original data. Both these tables are contained in the Appendix. The peer effects estimations have also been performed including all observations for which test scores are reported. Missing values have been replaced by zeros and dummy variables for missing values for each variable have been added to the regressions. The estimated peer effects are comparable to the evidence presented in section five and are available from the authors upon request.

The home questionnaire asked parents to report the number of books in their home in five categories: none or few books (0 – 10), enough to fill one shelf (11 – 25), enough to fill one bookcase (26 – 100), enough to fill two bookcases (101 – 200), enough to fill three or more bookcases (more than 200).[1] In order to control for an individual students' background, we use a set of dummy variables for these categories. In order to form a single peer measure, after some experimentation, we chose a simple index which assigns 1 to the lowest category (0 – 10), and 5 to the highest category (more than 200). The median parent reports 26 – 100 or 101 – 200 books, and the mean of the indices range from about 3.3 to 4, depending on the country (see Table 3 below and Table A1).

We generated peer variables as the class average of five student background variables: number of books at home, student's sex and age, whether at least one parent was born abroad, and whether a foreign language is spoken at home. There is an argument in the literature on peer effects whether class rooms or schools (or possibly even neighborhoods) are the more appropriate unit of peer interactions. Of course, peer interactions may occur at each of these levels, and it is an open question which is the most important. We focus on the class room level for the pragmatic reason that we want to analyze differences within schools. In the within school estimates, all peer interactions with students from other classes in the school will be absorbed into the school fixed effects. However, peer effects in the class room are clearly of interest, since classes are the basic unit where learning takes place. It is therefore natural to expect that a large fraction of total peer effects should arise at the class room level.

The peer averages are formed using information for all students that report a value for this specific variable in the data set, not just those students in the final sample. In Table 3, we decompose the total variance in these class averages into the parts of the variance within and between schools using the relationship

$$\frac{1}{N}\sum_{s=1}^{S}\sum_{c=1}^{C_s}\left(x_{cs}-\bar{x}\right)^2=\frac{1}{N}\sum_{s=1}^{S}\sum_{c=1}^{C_s}\left(x_{cs}-\bar{x}_s\right)^2+\frac{1}{N}\sum_{s=1}^{S}C_s\left(\bar{x}_s-\bar{x}\right)^2 \tag{4}$$

[1] Using instead the number of books at home reported by students yields comparable results, which are provided by the authors upon request.

where x is the specific variable we are interested in, $s = 1, 2, \ldots, S$ is a school indicator, $c = 1, 2, \ldots, C_S$ is a class indicator, and there are C_S classes in school s. N is the total number of classes across all schools in our sample.[2]

Table 3 presents the total, between and within school variance of the peer variables. The variation for the average reading test score is shown as well. It is obvious that most of the variance in all of these measures is between schools. Between 7 and 18 percent of the variance in the index for the number of books at home is within schools. The fraction is higher for the reading test scores. However, 70 percent or more of the test score variation is also between schools. This suggests that a large part of the variation in all these measures is accounted for by school effects. Nevertheless, there is also some non-negligible amount of variance left within schools.

4 Selection in class room formation

In this section we will discuss the assignment of students both between and within schools. We start by presenting some basic information on primary schooling in the countries we study. We then go on to present some evidence from the PIRLS data to shed light on the question whether classes are formed (more or less) randomly.

In all six countries in our sample, students attend a single track primary school from school enrolment to at least grade four, in which students have been tested in PIRLS.[3] While students are assigned to various school types after grade four in Germany, they stay on for at least two more years in primary school in most other countries (France, Iceland, the Netherlands, and Sweden) or go on to a single tracked secondary school (Norway). School choice for primary school is free in some countries (Germany and the Netherlands) and depends on the place of living in the other countries. However, parents have some means to influence the choice of schools also in these countries. In practice, most parents choose the nearest school for convenience in all countries. The heads of the school are responsible for the assignment of students to classes within schools. Most countries have legal rules on maximum class size and some school systems provide extra resources for schools with a high share of migrant

[2] For the variance decomposition to add to the total variance in an unbalanced panel, it is necessary to weight the between component by the number of classes in the sample. This is not what, for example, the Stata `xtsum` command calculates.

[3] The information on the schooling systems is taken from Eurybase, the database in the information network on education in Europe, http://www.eurydice.org.

students. The final responsibility lies with the heads of school, however. Grouping of students seems to happen in some cases based on the migration background of students. We have found no direct evidence that primary schools in any of the sample countries use any systematic ability grouping or sorting by family background in grade four.

We investigate two separate and distinct questions about class room formation with the PIRLS data. The first question is whether class rooms which differ in composition, for random or non-random reasons, receive different resources. The second question is whether the data are consistent with classes being formed randomly. In order to shed light on the first question, we ran a set of regressions of the peer variables described in the previous section on class room, teacher, and school characteristics. The observable characteristics of class rooms and schools which we use are class size, teacher gender, education, and experience, size of the town or city, instruction hours per day, and the degree of shortages of staff, materials, or rooms that the school reports.

Table 4 illustrates these regressions for Germany. For books at home, the only significant variable is shortage of rooms in the school, while none of the class level variables are individually significant. The other peer characteristics are also typically unrelated to class room and school characteristics, with coefficients usually being small and insignificant. The only exception is immigrants and non-native speakers, who tend to be clustered in more urban schools. Classes with more immigrants also tend to have better educated teachers.

Table 5 shows p-values for the corresponding F-tests on the joint significance of the variables. The first three rows of this table refer to simple OLS regressions of the type run in Table 4, which include both class room and school characteristics. F-tests for the class room level variables, the school level variables, and both groups together are reported. The fourth row reports the F-tests for the class level variables from a regression including school fixed effects. For our family background variable of interest, the number of books at home, the class and school variables are typically insignificant, with a few exceptions. School level variables seem to matter in the Netherlands and Norway. This may simply indicate that students from more favorable backgrounds are not randomly allocated to schools. There is no correlation with class level variables, however. So this should not affect our conclusions when we look at within school results. There is some evidence for class level variables being related to family background in Iceland and in Sweden. In the case of Iceland, this only appears in the fixed effects regressions. It turns out that this correlation is solely driven by a single class room with a teacher with 20 years of experience (while all other teachers in Iceland have 10 or fewer years of experience). We discount this result as spurious. In the case of Sweden this seems indeed to indicate a non-random allocation of class room resources to classes with students from different background, even within schools. There is evidence that class size increases with average background of students in a class. The coefficients for the other class and school variables are not significant.

We also find some evidence that class rooms differ for students by age (in Germany, Iceland, and Norway) and by student sex (in Iceland, the Netherlands, and Norway). It also seems fairly clear that classes are different for immigrant students in all our sample countries. The higher the share of immigrant students in a class, the lower is teacher's education in Germany and Norway. Instead, teacher's education increases with the share of immigrants in France. In Sweden, there is weak evidence for an allocation of immigrant students to larger classes.

In order to test whether class rooms are formed randomly with respect to a particular student characteristic, we perform a series of Pearson Chi² tests. If classes are formed randomly, the student characteristic under study and the class the student is assigned to should be statistically independent. For example, consider student sex. The Pearson Chi² test asks whether there are more females in a particular class than is consistent with independence, given the number of students in the school. Formally, for each school the test statistic is given by

$$P = \sum_c \sum_j \frac{\left(n_{cj} - \hat{n}_{cj}\right)^2}{\hat{n}_{cj}} \tag{5}$$

where n_{cj} is the number of students with characteristic j in class room c, $c = 1, \ldots, C$, $j = 1, \ldots, J$. Define

$$n_{c\bullet} = \sum_j n_{cj} \qquad n_{\bullet j} = \sum_c n_{cj} \qquad \hat{n}_{cj} = \frac{n_{c\bullet} n_{\bullet j}}{\sum_c \sum_j n_{cj}}$$

where $\hat{n}_{cj}$ is the predicted number of students with characteristic j in class room c when characteristic and class room are independent. Then, under the null hypothesis of independence, $P \sim \chi^2$ with $(C-1)(J-1)$ degrees of freedom.

We further assume that the S schools in a country are independent. In this case, we can simply add up the S test statistics to get an aggregate test statistic with $\left[\sum(C_s - 1)\right](J-1)$ degrees of freedom (see, e.g. DeGroot, 1984, p.384). Obviously, the test can only be carried out on the sub-sample of schools with two or more class rooms. These test statistics are shown in the fifth row of each panel in Table 5. We found in a small Monte Carlo experiment that the test generally performs well but rejects somewhat too often under the null in samples of our size.

Figure 1 shows the histograms of the school specific p-values for this test for Germany. Under the null hypothesis of random assignment to classrooms, the p-values should be (roughly) uniformly distributed. A policy of deliberate balancing of students' characteristics across classrooms (e.g. making sure that the same fraction of girls are in each class) would lead to a left-skewed distribution, while an assignment of similar students to the same classroom would lead to a right-skewed distribution (cf. Graham, 2004). Figure 1 suggests that students are roughly randomly assigned to classes on the basis of family background and sex. However, there is clustering of immigrant and non-native language students in certain classes, as low p-values predominate for these tests.

Table 5 confirms these results. The p-values of the aggregate tests for books at home and sex are 0.24 and 0.12, respectively. However, the p-values are only 0.016 for foreign born, and 0.050 for non-German speaking children. We also find evidence of non-random assignment of immigrant children for the Netherlands and Sweden. In addition, there seems to be some evidence that children of a similar family background are grouped together in Sweden.

The school questionnaire also provides information on the non-random allocation of students to classrooms. Principles are asked whether fourth-grade classes are formed on the basis of student ability in their school. The fraction of students in schools for which principles affirm this question is reported in the last row in Table 2 and equals around five percent in most countries. Only in France with 28 and The Netherlands with 33 percent the non-random allocation of students seems to be more common. The average test scores of students in schools that affirm to apply streaming hardly differ from other schools. Only in Norway students in schools that form classes on the basis of student ability score 12 test score points higher than students in other schools. In Iceland, there is only one school that says to apply streaming and which scores lower than the average. When we estimate the peer effects in the following section, students in the respective schools are excluded from the regression in one specification.

Our results largely confirm that classes in the sample countries seem to be formed roughly randomly within schools. There is little evidence that students of different family backgrounds are more likely to be grouped in certain classes conditional on the school they attend, or that classes with different compositions receive different (observable) resources. This is comforting for our analysis. The only country, where this does not seem to be the case, is Sweden. Hence, the Swedish results may have to be taken with a grain more of salt. In addition, immigrant children, which are an important group in all of the sample countries, also seem to be non-randomly assigned and given different teaching resources.

5 Results on peer effects

We now turn to our results on peer effects. Table 6 reports results for Germany, which is the country with the biggest sample, and, in particular, with the biggest number of schools with at least two classrooms. We report results for two versions of the peer variable on books at home. Since the variable takes on five categories, the most straightforward way to present the results is to use the fraction of peers in four of the categories. Alternatively, we use the 1 – 5 index we created from these five categories. In each case, the peer variable for student i used in the regressions is the leave-out mean for the classroom, omitting the value of the variable for student i from the calculation of the mean.

Column 1 shows results when we only control for student level characteristics (gender, age, immigrant status, language at home, household size, and four dummies for the number of books in the home of the student). Going from a classroom where all peers have 26-100 books to a classroom where all peers have 101-200 books, a change of about a two standard deviations of the peer composition across classrooms, is associated with a 19 point increase in the test score. This is about 32 percent of the student level standard deviation in test scores. We find a very similar result in the specification with the single index measure for peers' books at home. This is not surprising since in practice most of the variation in the index variable occurs in the range between the 26-100 and the 101-200 books categories.

In columns (2) and (3) of Table 6, we add class and school level covariates to the regression. Adding these covariates does little to the results. This is not surprising, since we found in Tables 4 and 5 above that the class and school level covariates are basically uncorrelated with our measure of peer quality. In column (4) we add school peer measures. This changes the results dramatically. The coefficient on the index variable drops from about 18 to zero and is insignificant. When the peer effects are measured by the fraction of peers in each category, the peer effects are no longer monotonic now. However, these effects are also imprecisely estimated. Finally, in columns (5) and (6), we add a full set of school fixed effects, leading to comparable results as for the school peer measures. The exclusion of schools that reported to form classes on the basis of student ability in column (6) hardly changes the results. The upshot from this exercise is that the peer effect for Germany falls to a relatively modest amount when we only look at variation in peer composition within schools.

Table 7 summarizes the results for all the six countries. To be more parsimonious, we only present the results using the index measure of peer quality. We find a relatively consistent pattern of results for all six countries in our sample. The size of the estimated peer effect is similar across the specifications with and without school and class level variables, and is in the order of 15 to 23 for moving peer quality to the next higher category. Only in Norway does the peer coefficient fall when school level covariates are added to the regression. Once we include school peer measures or school fixed effects, the effect always falls, although the amount of the change is different across countries. Only in France, the coefficient increases when school peer measures are added and falls only slightly with school fixed effects. In Norway and Germany the decrease is strongest. Excluding schools that form classes based on student ability only changes the results in France and The Netherlands. These are the two countries with high shares of students in schools that apply streaming (see Table 2). While the coefficient in France decreases, it increases in The Netherlands. We would expect the coefficient to decrease once schools that apply streaming are excluded. The result for The Netherlands is thus curious.

One reason for the high variation in the coefficients from the fixed effects models is that the standard errors of these estimates are reasonably large, so that the effects for each individual country cannot be estimated very precisely. If we believe that the peer effects are the same in each country then it makes sense to combine the estimates into a single estimate. The average of the six coefficients, weighted by the inverse of the sampling variance, is 8.3. If the variation in country level estimates around this overall mean is only due to sampling variation, then the standard error for the meta-estimate is 3.3.[4] This estimate is much more precise than the country level estimates, and it is significant at the 5 percent level. One concern is with the results for Sweden, because we found some evidence for non-random assignment and target class room resources for Sweden above. The meta-estimate for the countries without Sweden is 7.3 with a standard error of 3.8.

These estimates are roughly in line with the findings in much of the literature. The across school standard deviation of the class room mean of the index of books at home is 0.50 on average in the six countries. This suggests that a one standard deviation change in the peer measure leads to an effect on reading test scores of 4.14. The average standard deviation of reading test scores in the six countries is 63.52, so this is an effect size of 0.07 of a standard deviation. This is in the middle of the range of estimates found in many previous studies of peer effects, although some studies found basically zero effects (e.g. Angrist and Lang, 2004) and others larger effects (e.g. Hoxby, 2000).

Our results show that standard OLS estimates of the peer effect are generally biased upward substantially if the within school results are indeed reliable estimates of the true peer effect. This is true even after controlling for the typically available measures of class room and school characteristics.[5] One reason why even the fixed effects estimates may be biased is the presence of immigrant children. We showed above that immigrant children are often not randomly assigned to classes within schools, and the classes with many immigrant children may get different resources. Since immigrants in these countries tend to be of lower SES (the index for books at home is on average 3.15 for immigrant families in the six countries but

---

[4] The sampling variance of the mean is obtained as $v = \left[\sum v_c^{-1}\right]^{-1}$, where v_c is the sampling variance of the estimate for country c. One interpretation of this calculation is that the country average is the minimum distance estimate of the common peer effect across countries.

[5] Including measures from the school questionnaire on the share of economically disadvantaged students, the share of immigrant students or the share of students leaving before the end of the academic year as alternative school level variables did not change the results.

3.56 for non-immigrant families), part of the peer effect may be explained by the non-random allocation of immigrants.

In order to probe this, we reran the regressions in Table 7 including the fraction of foreign born children in the class, and the fraction of children speaking a foreign language at home. This attenuates the estimated peer effects at most very slightly. We also experimented with regressions on the sub-sample of schools with few immigrant children. However, most sample countries have enough immigrants that there are relatively few such schools leading to small samples, and hence imprecise estimates.

The peer effects have also been estimated for parents' education as a peer measure. Table A4 in the Appendix presents the results along with estimates for the number of books at home using the same sample. In the specification including only student level variables, the estimates for both peer measures are of comparable size. In the school fixed effects specification in column (2), the peer effects for parents' education are much lower than for the number of books at home, though.

A further question is whether peer effects vary across students. This could give insights into the optimal assignment of students to classes. When students from a lower social background profit more from their peers' background than students from a high social background, more heterogeneous classes would benefit overall performance (Glewwe, 1997). To investigate this, we add interaction effects between the peer variable and the individual variable books at home to the regressions presented in Table 7. Since about half the students have more than 100 books at home, we interact the peer average with a dummy indicating whether the individual reports more than 100 books at home. The results are presented in Table 8. We find no evidence that the peer effects are non-linear with respect to student background. Peer effects tend to be slightly stronger for students with a higher social background in France and the Netherlands, slightly weaker in Norway and Sweden, and zero in Germany and Iceland. However, the interaction terms are insignificant in each case. The same is true for the average effect across all countries.

6 Conclusion

Peer effects are a major input into the process of educational production but are difficult to estimate empirically. We estimate peer effects across classes within schools and argue that classes within schools are formed randomly with respect to family background. We find that a one standard deviation change in our student background measure of peer composition leads to a 0.07 standard deviation change in reading test scores of fourth graders across our sample of six European countries. This is fairly similar to previous estimates of peer effects. The estimated peer effects are highest in France and The Netherlands and lowest in Norway, Iceland and Germany. For Sweden, the estimated effects are also high but might be driven by within-school selection. We find no significant evidence on non-linear peer effects with respect to student background.

The results indicate that sorting across schools is strong and cannot be fully controlled for by most available measures of class and school characteristics. Hence, standard OLS estimates of peer effects that do not control for schools are generally biased upward. The small but significant peer effects imply that the composition of classrooms affects student learning. However, the evidence on non-linear peer effects and thus on the optimal mixing strategy of students is not clear-cut for the primary schools of the European countries investigated.

References

Ammermueller, Andreas (2005), Educational opportunities and the role of institutions, ZEW Discussion Paper No. 05-44.

Angrist, Joshua and Kevin Lang (2004), Does school intergration generate peer effects? Evidence from Boston's Metco Program, *American Economic Review* 94(5), 1613-1634.

Boozer, Michael and Stephen Cacciola (2001), Inside the 'Black Box' of Project STAR: Estimation of peer effects using experimental data, Economic Growth Center Center Discussion Paper No. 832, Yale University.

Cullen, Julie, Brian Jacob and Steven Levitt (2003), The effect of school choice on student outcomes: Evidence from randomized lotteries, NBER Working Paper No. 10113.

DeGroot, Morris (1984), *Probability and statistics*, 2nd edition, Reading, MA: Addison Wesley Publishing Company.

Gibbons, Stephen and Shqiponja Telhaj (2005), Peer effects and pupil attainment: Evidence from secondary school transition, London School of Economics, mimeo.

Glewwe, Paul (1997), Estimating the impact of peer group effects on socioeconomic outcomes: Does the distribution of peer group characteristics matter? *Economics of Education Review.*

Gould, Eric, Victor Lavy and Daniele Paserman (2005), Does immigration affect the long-term educational outcomes of natives? Quasi-experimental evidence, Hebrew University, mimeo.

Graham, Bryan (2004), Identifying social interactions through excess variance contrasts, Harvard University, mimeo.

Hanushek, Eric, John Kain, Jacob Markman and Steven Rivkin (2003), Does peer ability affect student achievement?, *Journal of Applied Econometrics* 18 (5), 527-544.

Hoxby, Caroline (2000), The effects of class size on student achievement: New evidence from population variation, *Quarterly Journal of Economics* 115(4), 1239-1285.

Manski, Charles (1993), Identification of endogenous social effects: The reflection problem, *The Review of Economic Studies* 60(3), 531-542.

McEwan, Patrick (2003), Peer effects on student achievement: Evidence from Chile, *Economics of Education Review* 22(2), 131-141.

Mislevy, Robert (1991), Randomization-based inference about latent variables from complex samples, *Psychometrika* 56, 177-196.

Mullis, Ina, Michael Martin, Eugene Gonzalez and Ann Kennedy, (2003), *PIRLS 2001 international report: IEA's study of reading literacy achievement in primary schools*, Chestnut Hill, Boston College.

Schindler-Rangvid, Beatrice (2003), Educational peer effects: Quantile regression evidence from denmark with PISA 2000 data, Kapitel 3 i "Do schools matter?", phd-afhandling nr. 2003:9 fra Handelshøjskolen i Århus.

Schneeweiss, Nicole and Rudolf Winter-Ebmer (2005), Peer effects in Austrian schools, Department of Economics, University of Linz, Working Paper No. 0502.

Woessmann, Ludger (2004), How equal are educational opportunities? Family background and student achievement in Europe and the United States, CESifo Working Paper No. 1162, Munich.

Figure 1: Distribution of school-specific p-values for Pearson Chi² test for Germany

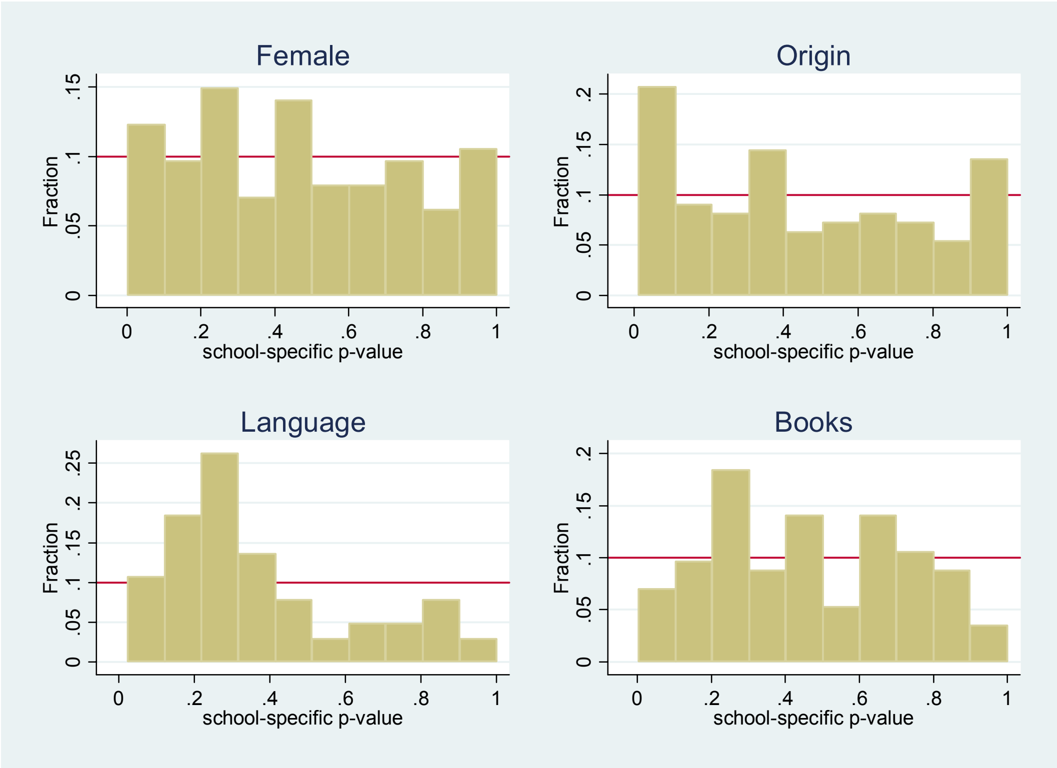

Table 1: Recent studies on peer effects

Studies	Context	Grade / student age	Outcome	Peer measure or treatment	Peer group	Identification	Magnitude of peer effect	Source of Magnitude	Non-linearities in peer effects
Hoxby (2000)	Public schools, Texas, US	3rd to 6th grade	Test scores (Reading, Math)	Test scores	Class	Cohort gender variation	1.s.d. → ~0.4 s.d.	Table 9, columns 3 and 4	No evidence
McEwan (2003)	Secondary schools, Chile	8th grade	Test scores (Spanish)	Mothers' education	Class	School fixed effects	1.s.d. → 0.27 s.d.	Table 3, column 4	Slightly concave
Hanushek et al. (2003)	Public schools, Texas, US	5th to 6th grade	Math test score gain	Math scores two years ago (also incl. prop. free lunch, std. dev of score)	School-grade	Student and school-by grade fixed effects	1.s.d. → 0.05.s.d.	Table 2, column 3 Table A1	No evidence
Cullen, Jacob and Levitt (2003)	Chicago public schools, US	9th and 10th grade	Test scores, and others	Winning a school voucher in lottery	-	Random assignment	Near zero and insignificant	Table 6, columns 1 and 2	No evidence
Schindler Rangvid (2004)	Secondary schools, Denmark	Age 15	PISA test scores (Reading)	Mothers' education	School-grade	Additional controls from register data	1 s.d. → 0.07 s.d.	Table 4 Table 1	Weak evidence for declining effects
Gibbons and Telhaj (2005)	State secondary schools, England	Age 14	Test scores (English, Math)	Attainment at age 11	School-grade	Random variation in transition from primary to secondary schools	1 s.d. → 0.08 s.d.	Tables 3 and 4, column 7	Slightly concave
Graham (2004)	STAR project, US	Kinder-garten to 3rd grade	Test scores (Math, Reading)	Test scores	Class	Random assignment and excess variance contrasts	50 percentile increase → 0.9-1.1 s.d.	Table 5, column 3	-
Schneeweis and Winter-Ebmer (2005)	Secondary schools, Austria	Age 15	PISA test scores (Reading)	Socio-economic status	School-grade	School type fixed effects	1 s.d. → 0.06 s.d.	Table 2, column 1 Table 1	Weak evidence for declining effects

Builds on Table 1 in Gibbons and Tehlhaj (2005).

Table 2: Mean scores and sample sizes
(Standard deviations in parentheses)

	Germany	France	Iceland	Netherlands	Norway	Sweden
Reading score (all)	539.09 (63.61)	525.17 (66.57)	512.42 (71.03)	554.21 (51.15)	499.18 (77.52)	561.01 (61.54)
Reading score (sample)	548.63 (59.91)	533.84 (64.31)	518.59 (68.42)	565.15 (51.25)	504.96 (76.04)	563.45 (61.47)
Reading score (excl. streamed schools)	548.60 (59.73)	534.55 (65.45)	518.68 (68.34)	562.84 (53.62)	504.44 (76.44)	563.13 (61.68)
Reading scores (streamed schools)	549.37 (64.24)	531.68 (60.70)	506.84 (81.04)	568.77 (47.46)	529.08 (66.51)	569.97 (56.62)
No. of students (all)	7,633	3,538	3,676	4,112	3,459	6,044
No. of students (sample)	4,577	2,267	1,728	1,857	2,548	3,940
No. of students in schools with > 1 class	3,628	1,583	1,301	805	1,748	3,225
No. of schools	183	115	84	105	117	119
No. of schools with > 1 class	114	55	39	29	54	79
No. of classes	301	172	135	141	171	267
No. of classes in schools with > 1 class	232	112	90	65	108	227
Fraction of students in schools that say to apply streaming	0.0671	0.2757	0.0064	0.3282	0.0352	0.0470

Scores are weighted by students' sampling probability, standard deviations are in parentheses. Rows 4 and below refer to the sample used in the estimations. Last row reports the fraction of students in schools in which principles state that classes are formed by ability of all students in schools for which principles reply to the question.

Table 3: Decomposition of variance in class level means

	Germany	France	Iceland	Netherlands	Norway	Sweden
			Index of the number of books at home			
Mean	3.45	3.30	4.00	3.37	4.00	3.95
Total	.2401	.3138	.1480	.3922	.1542	.2643
Between	.2098	.2726	.1220	.3629	.1297	.2174
Within	.0303	.0412	.0259	.0293	.0245	.0469
			Age			
Total	.0326	.0313	.0065	.0306	.0082	.0111
Between	.0250	.0183	.0050	.0212	.0060	.0060
Within	.0076	.0130	.0015	.0094	.0022	.0051
			Female			
Total	.0145	.0226	.0212	.0156	.0145	.0158
Between	.0085	.0174	.0170	.0139	.0125	.0091
Within	.0061	.0052	.0043	.0017	.0020	.0067
			Foreign parent			
Total	.0459	.0463	.0095	.0488	.0222	.0485
Between	.0404	.0413	.0069	.0451	.0189	.0386
Within	.0054	.0050	.0026	.0036	.0033	.0099
			Foreign language at home			
Total	.0141	.0151	.0088	.0345	.0069	.0230
Between	.0112	.0128	.0058	.0330	.0052	.0167
Within	.0029	.0023	.0030	.0015	.0017	.0064
			Reading test scores			
Total	1144.71	1223.61	751.93	896.62	1075.93	1123.78
Between	978.47	908.63	569.62	799.28	933.10	791.51
Within	166.24	314.97	182.31	97.34	142.83	332.27

Table 4: Regressions of peer variables on class room and school characteristics in Germany

Independent Variable	Dependent Variable				
	Books	Age	Female	Foreign parent	Foreign language
Class size	.002	-.011	.030	.010	.015
	(.091)	(.043)	(.017)	(.028)	(.014)
Class size squared*100	.03	.001	-.06	-.03	-.03
	(.2)	(.09)	(.04)	(.06)	(.03)
Female teacher	.109	.031	.006	-.008	-.009
	(.070)	(.024)	(.019)	(.030)	(.013)
Teacher education:	.001	.026	-.039	.006	-.018
ISCED 4	(.146)	(.050)	(.039)	(.036)	(.023)
Teacher education:	.065	-.062	-.023	.173	.056
ISCED 5+	(.114)	(.047)	(.034)	(.025)	(.015)
Teacher's experience	.001	-.002	-.005	.005	.0001
in grade 4	(.011)	(.003)	(.003)	(.004)	(.002)
Teacher's exp.	.003	.008	.01	-.01	.00002
Squared*100	(.03)	(.009)	(.008)	(.01)	(.006)
Size of town	-.077	.141	.049	-.084	-.059
< 3000 inhab.	(.265)	(.037)	(.074)	(.048)	(.039)
Size of town	.008	.086	-.029	.090	.044
< 100,000 inhab.	(.081)	(.026)	(.020)	(.030)	(.014)
Size of town	.045	.073	.004	.167	.084
< 500,000 inhab.	(.143)	(.047)	(.031)	(.044)	(.028)
Size of town	-.046	.115	.017	.230	.116
> 500,000 inhab.	(.154)	(.049)	(.035)	(.064)	(.033)
Instruction hours per	.018	.005	-.022	.014	.012
day	(.053)	(.018)	(.012)	(.022)	(.013)
Shortage of staff	-.020	.004	-.009	.034	.011
	(.055)	(.018)	(.010)	(.021)	(.011)
Shortage of material	.005	-.035	-.001	-.013	.005
	(.071)	(.022)	(.013)	(.025)	(.015)
Shortage of rooms	-.112	.041	.010	.004	.001
	(.057)	(.022)	(.012)	(.020)	(.013)

Regressions are weighted by the students' sampling probability. Number of observations is 4,577. Standard errors are clustered at the school level.

Table 5: P-values for F-tests on class room and school characteristics and for Chi² test for independence of peer variable and class assignment

	Germany	France	Iceland	Netherlands	Norway	Sweden
			Index of the number of books at home			
Class	0.4532	0.2477	0.7094	0.4051	0.0995	0.0629
School	0.7095	0.1222	0.0595	0.0005	0.0431	0.4069
Class+school	0.7766	0.0554	0.3538	0.0000	0.0286	0.2857
Class (SFE)	0.4595	0.2647	0.0123	0.3370	0.5675	0.0000
Pearson Chi²	0.2415	0.4708	0.7964	0.7512	0.0893	0.0479
			Age			
Class	0.0002	0.4605	0.0704	0.0249	0.1587	0.3891
School	0.0050	0.0122	0.3197	0.0015	0.4364	0.1769
Class+school	0.0000	0.0341	0.1603	0.0000	0.2139	0.2642
Class (SFE)	0.0017	0.2482	0.0021	0.0046	0.0000	0.8663
Pearson Chi²	0.0694	0.3208	0.1452	0.0992	0.0467	0.5847
			Female			
Class	0.1270	0.7678	0.0006	0.0286	0.0089	0.9706
School	0.5170	0.3311	0.5958	0.0000	0.8831	0.3005
Class+school	0.4082	0.5105	0.0062	0.0000	0.0902	0.6486
Class (SFE)	0.5677	0.2751	0.0000	0.1467	0.0036	0.0909
Pearson Chi²	0.1240	0.4340	0.9608	0.6011	0.8827	0.9608
			Foreign parent			
Class	0.0000	0.1341	0.1339	0.0254	0.5828	0.3449
School	0.0000	0.0001	0.7760	0.0000	0.0000	0.0000
Class+school	0.0000	0.0005	0.3009	0.0000	0.0000	0.0000
Class (SFE)	0.0003	0.0306	0.0000	0.0024	0.0000	0.0054
Pearson Chi²	0.0157	0.3270	0.3201	0.0444	0.1335	0.0030
			Foreign language at home			
Class	0.0052	0.5956	0.0173	0.1830	0.3365	0.2167
School	0.0004	0.0085	0.6573	0.1222	0.1026	0.0000
Class+school	0.0000	0.0588	0.0476	0.0000	0.0867	0.0000
Class (SFE)	0.0000	0.3776	0.0001	0.0000	0.0029	0.0001
Pearson Chi²	0.0495	0.7505	0.1861	0.4217	0.4860	0.0010

The first four rows in each panel report p-values of Wald tests for the joint significance of the respective group of variables. The first three rows refer to regressions of peer variables on class and school variables. The fourth row refers to regressions of peer variables on class variables controlling for school fixed effects. Running the regressions at the class level or including the student level variables sex, age, origin and language in the regressions for the first three rows yield comparable results. The fifth row reports the p-value for Pearson Chi² tests of independence between the peer characteristic and class room assignment within each school using the individual level data. See text for details.

Table 6: Regressions for reading test score on peer composition for Germany
(Standard errors in parentheses)

Independent Variable	(1)	(2)	(3)	(4)	(5)	(6)
Peers' index of books at home	17.93	17.72	17.62	-.93	5.92	6.82
	(3.04)	(3.03)	(3.03)	(6.03)	(6.15)	(6.30)
Fraction peers 11-25 books at home	-20.15	-6.52	-4.42	-31.63	-49.91	-68.77
	(23.44)	(21.59)	(21.11)	(22.87)	(42.75)	(44.67)
Fraction peers 26-100 books at home	21.41	30.07	28.10	-17.32	-22.90	-26.36
	(18.08)	(17.01)	(16.09)	(21.24)	(32.38)	(33.20)
Fraction peers 101-200 books at home	40.48	49.32	47.49	-10.31	-11.15	-16.21
	(18.88)	(17.49)	(17.24)	(24.85)	(39.44)	(40.55)
Fraction peers 200+ books at home	54.86	61.87	62.18	-28.67	-15.95	-21.31
	(17.69)	(16.60)	(16.16)	(30.35)	(36.86)	(37.36)
Student level variables	✓	✓	✓	✓	✓	✓
Class level variables		✓	✓	✓	✓	✓
School level variables			✓	✓		
School peer variables				✓		
School fixed effects					✓	✓
Exclude streamed schools						✓
R^2	0.256	0.261	0.266	0.266	0.301	0.301

Weighted least squares regressions using students' sampling probability. Number of observations is 4,577. Standard errors are robust to clustering at the school level. Student level variables are student's sex and age, parents' origin, language spoken at home, number of books at home and number of persons living in household. Class level variables are class size, class size squared, teacher's sex, education, experience and experience squared. School level variables are community size, average daily instruction hours, shortage of staff, teaching material and buildings. School peer variables are school averages of index of number of books at home and parents' origin. Streamed schools are those for which principles state that fourth-grade classes are formed on the basis of ability. The R^2 from regressions where the peer variable is an index of books at home is reported.

Table 7: Regressions for reading test score on peer composition
(Standard errors in parentheses)

	(1)	(2)	(3)	(4)	(5)	(6)
Germany	17.93 (3.04)	17.72 (3.03)	17.62 (3.03)	-.93 (6.03)	5.92 (6.15)	6.82 (6.30)
France	21.84 (3.06)	22.81 (3.00)	22.21 (2.95)	23.59 (8.02)	18.83 (8.50)	12.59 (10.76)
Iceland	18.11 (5.79)	18.68 (5.46)	20.02 (5.07)	7.19 (12.16)	7.74 (11.68)	4.40 (11.72)
The Netherlands	17.49 (4.27)	16.99 (4.21)	19.67 (4.33)	6.09 (8.63)	10.65 (9.35)	20.07 (10.59)
Norway	15.11 (7.30)	14.33 (7.35)	9.46 (7.34)	3.50 (11.36)	-7.25 (9.44)	-6.90 (9.58)
Sweden	19.08 (3.90)	18.59 (3.87)	17.76 (4.17)	8.44 (6.44)	11.68 (7.13)	11.38 (7.37)
Student level variables	✓	✓	✓	✓	✓	✓
Class level variables		✓	✓	✓	✓	✓
School level variables			✓	✓		
School peer variables				✓		
School fixed effects					✓	✓
Exclude streamed schools						✓

Weighted least squares regressions using students' sampling probability. Peers' index of books at home is independent variable. Standard errors are robust to clustering at the school level. Student level variables are student's sex and age, parents' origin, language spoken at home, number of books at home and number of persons living in household. Class level variables are class size, class size squared, teacher's sex, education, experience and experience squared. School level variables are community size, average daily instruction hours, shortage of staff, teaching material and buildings. School peer variables are school averages of index of number of books at home and parents' origin. Streamed schools are those for which principles state that fourth-grade classes are formed on the basis of ability.

Table 8: Regressions for reading test score on peer composition and interactions with individual family background

(Standard errors in parentheses)

	(1)		(2)		(3)		(4)	
	Peer	Inter	Peer	Inter	Peer	Inter	Peer	Inter
Germany	16.62	3.02	-2.39	4.87	6.20	-.83	6.86	-.11
	(3.99)	(4.77)	(5.95)	(4.16)	(6.27)	(4.71)	(6.44)	(4.87)
France	20.76	2.73	21.16	4.80	16.58	4.23	11.95	2.31
	(3.85)	(4.62)	(8.14)	(4.44)	(8.65)	(4.65)	(11.27)	(5.49)
Iceland	20.99	-4.35	10.50	-4.90	8.39	-.95	4.76	-.53
	(8.83)	(9.82)	(14.19)	(10.73)	(15.04)	(10.73)	(15.08)	(10.80)
The Netherlands	16.60	2.13	3.36	5.99	7.65	6.29	16.32	8.12
	(5.30)	(6.56)	(9.37)	(6.17)	(10.33)	(6.41)	(10.87)	(7.06)
Norway	17.13	-2.65	9.91	-8.67	-2.29	-6.75	-2.43	-6.06
	(11.81)	(12.58)	(16.33)	(13.18)	(14.65)	(13.19)	(15.03)	(13.64)
Sweden	17.81	1.69	5.17	4.59	13.82	-2.76	14.69	-4.43
	(6.15)	(6.79)	(7.93)	(7.22)	(7.50)	(5.94)	(7.87)	(6.35)
Student level variables	✓	✓	✓	✓	✓	✓	✓	✓
Class level variables			✓	✓	✓	✓	✓	✓
School level variables			✓	✓				
School peer variables			✓	✓				
School fixed effects					✓	✓	✓	✓
Exclude streamed schools							✓	✓

Weighted least squares regressions using students' sampling probability. Peers' index of books at home (columns "Peer") and interaction term of peers' index and individual level dummy variable for > 100 books at home (columns "Inter") are jointly included. Standard errors are robust to clustering at the school level. Student level variables are student's sex and age, parents' origin, language spoken at home, number of books at home and number of persons living in household. Class level variables are class size, class size squared, teacher's sex, education, experience and experience squared. School level variables are community size, average daily instruction hours, shortage of staff, teaching material and buildings. School peer variables are school averages of index of number of books at home and parents' origin. Streamed schools are those for which principles state that fourth-grade classes are formed on the basis of ability.

Appendix

Table A1: Weighted mean values
(standard deviations in parentheses)

	Germany	France	Iceland	Netherlands	Norway	Sweden
Student background						
Student female	.52 (.50)	0.49 (0.50)	0.50 (0.50)	0.53 (0.50)	0.49 (0.50)	0.49 (0.50)
Student's age	10.51 (.47)	10.09 (0.49)	9.72 (0.28)	10.24 (0.46)	9.97 (0.33)	10.80 (0.32)
Foreign parent	.22 (.42)	0.29 (0.46)	0.12 (0.32)	0.17 (0.37)	0.17 (0.38)	0.23 (0.43)
Foreign language at home	.09 (.29)	0.12 (0.33)	0.11 (0.32)	0.13 (0.34)	0.08 (0.26)	0.09 (0.29)
Books at home (index 1 – 5)	3.49 (1.19)	3.32 (1.25)	3.99 (1.01)	3.36 (1.29)	4.03 (1.04)	3.92 (1.11)
0 – 10 books (ref.)	.06 (.24)	0.10 (0.30)	0.02 (0.13)	0.10 (0.31)	0.02 (0.13)	0.04 (0.19)
11 – 20 books	.12 (.32)	0.14 (0.35)	0.05 (0.21)	0.15 (0.35)	0.05 (0.22)	0.07 (0.25)
21 – 100 books	.36 (.48)	0.35 (0.48)	0.27 (0.44)	0.30 (0.46)	0.26 (0.44)	0.26 (0.44)
101 – 200 books	.19 (.39)	0.18 (0.38)	0.26 (0.44)	0.19 (0.39)	0.23 (0.42)	0.23 (0.42)
More than 200 books	.27 (.44)	0.24 (0.43)	0.41 (0.49)	0.26 (0.44)	0.44 (0.50)	0.41 (0.49)
Household size	3.54 (1.50)	3.63 (1.45)	3.78 (1.35)	3.51 (1.15)	3.49 (1.17)	3.56 (1.26)
Class						
Class size	22.77 (3.93)	24.04 (3.41)	19.56 (3.98)	26.65 (5.26)	20.32 (5.40)	24.20 (6.26)
Teacher female	.80 (.40)	0.62 (0.48)	0.94 (0.23)	0.51 (0.50)	0.84 (0.36)	0.78 (0.41)
Teacher's education ISCED 3 (ref.)	.07 (.25)	0.43 (0.49)	0.13 (0.34)	1.00 (0.00)	0.02 (0.14)	1.00 (0.00)
Teacher's education ISCED 4	.01 (.12)	0.23 (0.42)	0.12 (0.33)	0	0.98 (0.14)	0
Teacher's education ISCED 5+	.92 (.28)	0.34 (0.48)	0.75 (0.43)	0	0	0
Teacher's experience	8.26 (7.25)	9.15 (8.90)	2.84 (2.73)	6.25 (5.81)	3.17 (2.35)	6.72 (6.49)

Table A1, continued: Weighted mean values
(standard deviations in parentheses)

	Germany	France	Iceland	Netherlands	Norway	Sweden
School						
Village (ref.)	.31	0.25	0.06	0.01	0.25	0.14
	(0.46)	(0.44)	(0.23)	(0.09)	(0.43)	(0.35)
< 3000 inhab.	.02	0.06	0.17	0.14	0.15	0.08
	(0.12)	(0.24)	(0.37)	(0.35)	(0.36)	(0.27)
< 100,000 inhab.	.47	0.51	0.45	0.68	0.43	0.64
	(0.50)	(0.50)	(0.50)	(0.47)	(0.50)	(0.48)
< 500,000 inhab.	.12	0.15	0.32	0.13	0.10	0.08
	(0.32)	(0.36)	(0.47)	(0.33)	(0.30)	(0.26)
> 500,000 inhab.	.09	0.02	0	0.04	0.07	0.07
	(0.28)	(0.13)		(0.20)	(0.25)	(0.26)
Instruction hours/day	3.76	5.06	4.21	5.02	3.47	4.38
	(0.65)	(0.35)	(0.54)	(0.16)	(0.77)	(0.76)
Shortage of staff	1.76	1.23	1.56	1.55	1.53	2.10
(index 1 – 4)	(0.81)	(0.45)	(0.81)	(0.71)	(0.91)	(1.04)
Shortage of material	1.30	1.50	1.70	1.51	1.53	1.53
(index 1 – 4)	(0.51)	(0.75)	(0.87)	(0.66)	(0.74)	(0.82)
Shortage of rooms	1.49	1.40	1.62	1.49	1.67	1.75
(index 1 – 4)	(0.69)	(0.76)	(0.94)	(0.76)	(0.86)	(0.93)

Mean values are weighted by the students' sampling probability.

Table A2: Percentage of missing values

	Germany	France	Iceland	Netherlands	Norway	Sweden
Reading score	0	0	0	0	0	0
Student female	0	0	0	0	0	0
Student's age	0.03	0.17	0.03	0.54	1.07	0.02
Foreign parent	9.60	7.80	6.80	2.48	2.46	3.16
Foreign lang. at home	4.40	3.00	3.51	1.53	2.20	2.38
Books	13.32	10.97	16.89	35.41	9.42	9.33
Siblings	8.23	6.53	4.68	4.60	3.79	4.42
Class size	6.86	6.16	6.86	11.84	1.24	4.27
Teacher female	8.14	6.50	7.64	11.31	0.69	4.24
Teacher's education	9.45	5.54	11.29	-	0.69	-
Teacher's experience	14.45	6.16	12.40	14.42	3.32	7.63
Size of town	4.48	6.73	23.91	11.87	2.66	4.33
Instruction hours/day	7.91	6.02	18.47	10.00	6.27	11.15
Shortage of staff	4.48	7.55	20.08	10.34	2.11	5.15
Shortage of material	5.04	5.17	21.33	10.80	2.60	4.91
Shortage of rooms	4.48	6.22	20.13	10.09	2.89	3.87

Table A3: Weighted means of student background variables by the number of books

	Germany				France			
	0-20 books	21-100	101-200	> 200 books	0-20 books	21-100	101-200	> 200 books
Student female	.49	.55	.51	.50	.50	.49	.50	.49
Student's age	10.68	10.52	10.45	10.42	10.26	10.10	10.03	9.97
Foreign parent	.40	.23	.15	.14	.34	.29	.24	.29
Foreign language	.19	.09	.06	.05	.21	.11	.11	.07
Household size	3.60	3.51	3.49	3.57	3.85	3.59	3.46	3.58
No secondary	.07	.02	.01	.01	.19	.09	.03	.03
Lower secondary	.29	.13	.05	.03	.52	.37	.20	.08
Upper secondary	.23	.27	.24	.20	.17	.24	.27	.12
Post secondary	.36	.46	.41	.23	.09	.16	.21	.20
Tertiary	.05	.12	.30	.54	.04	.13	.28	.57
Income < 30,000$	.63	.45	.30	.17	-	-	-	-
30,000$ – 50,000$	.32	.41	.47	.43	-	-	-	-
> 50,000$	.06	.15	.23	.41	-	-	-	-
Number of obs.	835	1,641	849	1,252	519	764	410	574

	Iceland				The Netherlands			
	0-20 books	21-100	101-200	> 200 books	0-20 books	21-100	101-200	> 200 books
Student female	.44	.50	.48	.53	.53	.54	.53	.51
Student's age	9.76	9.71	9.74	9.71	10.33	10.23	10.19	10.17
Foreign parent	.25	.12	.10	.10	.25	.17	.11	.12
Foreign language	.15	.12	.11	.11	.18	.15	.13	.09
Household size	3.67	3.76	3.74	3.84	3.41	3.47	3.47	3.68
No secondary	.06	.02	.01	.01	.02	.01	.002	.002
Lower secondary	.22	.18	.16	.08	.69	.56	.36	.15
Upper secondary	.55	.54	.47	.27	.14	.18	.13	.09
Post secondary	.12	.11	.09	.11	.10	.22	.40	.45
Tertiary	.05	.15	.28	.53	.05	.03	.11	.31
Income < 30,000$	.34	.27	.20	.15	.48	.34	.25	.16
30,000$ – 50,000$	.49	.43	.41	.32	.37	.41	.39	.33
> 50,000$	.17	.30	.39	.54	.16	.25	.35	.51
Number of obs.	110	467	440	711	448	562	336	511

Table A3, continued: Weighted means of student background variables by the number of books

	Norway				Sweden			
	0-20 books	21-100	101-200	> 200 books	0-20 books	21-100	101-200	> 200 books
Student female	.47	.50	.45	.50	.49	.47	.52	.50
Student's age	9.99	9.97	9.97	9.97	10.80	10.81	10.80	10.80
Foreign parent	.31	.17	.14	.16	.45	.26	.19	.17
Foreign language	.21	.10	.06	.05	.23	.11	.07	.05
Household size	3.46	3.37	3.42	3.60	3.85	3.57	3.50	3.51
No secondary	.01	.004	0	.001	.03	.003	.002	.001
Lower secondary	.09	.05	.02	.01	.17	.08	.05	.02
Upper secondary	.73	.64	.47	.26	.52	.50	.39	.22
Post secondary	0	0	0	0	.16	.24	.30	.21
Tertiary	.18	.30	.51	.73	.12	.17	.26	.54
Income < 30,000$	.45	.28	.19	.12	.46	.30	.21	.14
30,000$ – 50,000$	.42	.47	.41	.31	.44	.51	.47	.38
> 50,000$	.13	.24	.41	.57	.10	.19	.32	.48
Number of obs.	178	670	565	1,135	346	953	851	1,790

Mean values are weighted by the students' sampling probability. Number of observations refers to variables included in Table A1. For the variables highest educational level attained by at least one parent and annual household income in US$ before taxes the number of observations is lower in most countries.

Table A4: Regressions for reading test score on alternative peer composition
(Standard errors in parentheses)

	(1)		(2)	
	Books	Parent's education	Books	Parent's education
Germany	18.79 (3.42)	14.76 (3.85)	9.39 (7.69)	5.17 (5.84)
France	22.41 (3.28)	17.96 (2.94)	24.29 (8.72)	6.98 (8.86)
Iceland	17.95 (5.89)	14.79 (3.44)	7.91 (11.83)	-6.82 (12.47)
The Netherlands	18.02 (4.77)	17.75 (3.86)	7.55 (10.21)	-30.15 (12.89)
Norway	15.21 (7.33)	15.06 (5.26)	-4.93 (8.62)	-9.00 (8.37)
Sweden	19.36 (3.93)	27.17 (4.49)	11.84 (7.11)	-.05 (7.60)
Student level variables	✓	✓	✓	✓
Class level variables			✓	✓
School fixed effects			✓	✓

Weighted least squares regressions using students' sampling probability. Peers' index of books at home and peers' index of parents' highest education are independent variables. Standard errors are robust to clustering at the school level. Student level variables are student's sex and age, parents' origin, language spoken at home, number of books at home and number of persons living in household. Class level variables are class size, class size squared, teacher's sex, education, experience and experience squared.