Should We Track or Should We Mix Them?

April 30, 2014

Abstract

In this paper, I propose and implement an innovative strategy to estimate peer effects in education. I describe the strategy and find that peer effects in the classroom are non-linear. I show that there are complicated interactions between classmates with different achievement levels. The findings are important for consideration of the classroom design and contribute to the discussion of tracking by abilities in schools. I conduct a hypothetical policy experiment to demonstrate the effects of shifting class composition and estimate gains and loses for students with different initial abilities.

KEYWORDS: peer interactions, ability grouping, achievement, instrumental variables.

Introduction

The recent surge in the study of peer effects is the logical consequence of policy debates about educational reforms. Indeed, the existence of knowledge spillovers in educational setting would allow for a wide range of policy interventions with the goal of building more efficient and equal classrooms, for example. Models of school choice implicitly assume the existence of peer effects and evaluate the consequences of the school choice program based on this premise (Epple and Romano 1998, Epple, Newlon and Romano 2002). Parents have been shown willing to pay for having good peers for their children in schools (Rothstein 2006). But is it true that only good peers matter? What do we know about the complementarity and substitutability of students with different abilities? In this paper, I propose and implement a new strategy to estimate peer effects in the classroom and investigate the non-linearities in classroom interactions. The findings allow me to evaluate the pros and cons of various classroom designs based on the diversity of abilities among students.

The interest in the non-linearities of peer effects in education is related to the ongoing debate about the benefits of ability tracking, or streaming, in schools or within the classroom.The standard argument in favour of tracking is that it is easier to teach a group with small variance of abilities. In that sense, the streaming of students by schools or classes is an efficient way to organize education process as teachers can specialize. At the same time, the opponents of tracking argue that grouping students by abilities prevents low ability students from benefiting from their high achieving peers. School tracking has been shown to increase inequality of opportunity and also to be detrimental to skill formation (Ammermuller 2005, Hanushek and Wößmann 2006). These studies also noted that another potential concern with tracking is that less effective teachers are assigned to classes with the majority of low ability students and less resources are directed towards such classes or schools.

Tracking as an institutional setting that characterizes compulsory education is debatable because, unlike sorting on family background (which also leads to increase in inequality), school design is relatively easily amendable by policy. For instance, England and countries in

Northern Europe have experimented over the last sixty years with school design, switching between fully comprehensive and segmented education. Now, in the majority of European countries, tracking takes the form of well-defined segregated tracks, while in the Unites States, tracking is represented by ability grouping within a fully comprehensive school structure. In Canada - where the data for this study come from - compulsory education is similar to the US and is comprehensive, with individual schools and principals making decisions about streaming for all or some subjects.

The implications of ability grouping, school choice programs and segregation depend on the nature and structure of peer effects. The benefits of tracking would arise if students from different backgrounds and ability levels experience peer effects of different magnitude from different peers. Since the theoretical models of peer interactions that evaluate the school choice programs as a rule assume average effect of peers, there is clearly a need to empirically estimate the consistency of this assumption with the data. A handful of studies that go beyond estimation of linear peer effects in schools concluded that data does not support the linear-in-means model of classroom interactions and found evidence of monotonicity of these effects (Hoxby and Weingarth 2006, Imberman et at 2012).

In this paper, I rely on the entry of new students to a class as a plausibly exogenous shift in the composition of the classmates and find that linear-in-means model hides important heterogeneity in the response of different students to different peers. Thus, looking at the shares of students with different initial ability, I find that the average peer effect overestimates the negative impact of low ability classmates on own achievement and underestimates the positive effect of high ability peers. Moreover, students from the different segments of the ability distribution experience a different impact from their peers, both in terms of the average quality of classmates and in terms of peers from the low or high ends of ability distribution. It turns out that independent of own ability, all students benefit when surrounded by good peers. While peer effect is achievement specific, the diversity of abilities in the classroom does not seem to be a factor that determines own achievement gain of a student.

What kind of implications do the findings of this study provide for policy-makers? First, there exist benefits of tracking for high ability students who unambiguously gain when grouped with similar well-achieving classmates. The implications for other students are less clear-cut. For instance, low achievers are better off when mixed with their high ability peers than they are when grouped with the same ability classmates. The net effect of tracking depends on the original distribution of abilities among students and whether the gain for low-achievers would overweigh the lost benefits for high ability students.

Previous Literature

In economics literature, the debate about school tracking is a debate about the trade-off between equality and efficiency in education.[1] The arguments about ability grouping in schools in educational literature are as a rule centered around issues of fairness and equality.[2] Both disciplines consider peer effects to be one of the reasons why ability tracking in schools may or may not be beneficial for school efficiency and for the reduction of inequality.[3]

There is a great divide between the opponents and supporters of school tracking. Broadly speaking, school tracking is defined as ability grouping with or without design of the specific curriculum for different ability groups. The opponents of tracking argue that channeling students into different tracks increases the inequality of opportunity and aggravates future economic inequality. Proponents of tracking usually cite the increased efficiency when students are grouped by abilities in schools or classes.

The two most commonly asked questions in the early tracking literature were: (1) whether tracking improves school efficiency, and (2) whether tracking leads to income inequality. These studies used small samples of schools in the United States or England and found small effects of tracking on achievement, but they did not account for the endogeneity of ability tracking in schools.[1] Later research used nationally representative longitudinal data sets and found a small but robust positive effect for students in high-ability track and a

[1]The summary of the earlier research can be found in the meta-studies by Slavin (1987, 1990).

negative effect for students in the low-ability track (Gamoran and Mare 1989, Argys, Rees and Brewer 1996, Hoffer 1992). These studies also found that tracking increased inequality. The most recent research on tracking in US schools uses modern econometric techniques to evaluate the effects of tracking on achievement and inequality. Applying selectivity correction methods, propensity score matching and instrumental variables to account for endogeneity of tracking in schools, Betts and Shkolnik (2000) and Figlio and Page (2002) do not find a significant effect of tracking on achievement and argue that previously found a negative effect on inequality is likely overestimated. Betts and Shkolnik (2000) argue that the absence of compelling evidence of the effect of tracking on achievement is due to the severe omitted variables bias and endogeneity problem in quantitative studies.

Unlike studies that compare schools with and without tracking within the same jurisdiction, a large body of literature takes a different approach and instead exploits cross-country or within-country differences in school policies related to tracking or the variation in the timing when school tracking starts. These studies estimate the effect of ability tracking on returns to education (Meghir and Palme (2005) for Sweden), income mobility (Pekkarinen, Uusitalo and Pekkala (2006) for Finland), and inequality (Ammermuller (2005) for England). Cross-country studies are often aimed at understanding whether school tracking exacerbates the role of parental background on inequality and outcomes. For instance, Ammermuller (2005) and Hanushek and Wößmann (2006) find that tracking generates educational inequality, but has no effect on average achievement. Brunello and Checchi (2007) find that relationship between parental background and educational outcomes of children are being reinforced in the presence of tracking.

One of the rare experimental evidence of the benefits of tracking is presented in Duflo, Dupas and Kremer (2008). They found that tracking in Kenyan elementary schools increased school efficiency – test scores in tracked schools were on average higher. Their results also suggest that tracking is beneficial for students in all ability quartiles with higher ability students gaining slightly more.

The distributional consequences of school choice programs such as vouchers and charter schools are also evaluated through the prism of peer effects when the school quality is determined by the average ability of the classmates. Thus, Epple and Romano (1998) raised the concern that school initiatives aimed at improvement in the quality of public education at the same time may diminish the quality of peers and harm students' achievement. A theoretical model of tracking in private and public schools predicts gains for high ability students and losses for low ability students in equilibrium. Tracking, however, would improve public school quality and increase average achievement (Epple, Newlon and Romano 2002).

In general, the implications of ability tracking in schools depend on the underlying nature of peer effects and peer interactions. Just knowing the average effect of peers independent of own ability and independent of the composition of peers is not enough to predict the effect of tracking on achievement of an individual student. In the next section I present a brief review of the literature that analyzed the non-linearities in peer interactions.

Non-linearities of peer effects in previous studies

Most of the researchers working on peer effects in education have by now agreed that the impact of the average quality of peers on individual achievement, or so-called linear effect, should not be the main parameter of interest. As opposed to linear effects, heterogenous peer effects that depend on both individual ability and peer ability are also referred to as non-linearities to distinguish them from the standard linear-in-means model. From the policy perspective, heterogeneous effects are more important, as these allow designing the classroom makeup that would benefit children with different abilities. So far, a handful of papers discussed and estimated heterogeneous peer effects in elementary education. The seminal paper by Hoxby and Weingarth (2006) popularized the two-way interaction model of peer effects. The study also provides the most detailed description of models of peer effects and tests these models using the fully saturated specification with 100 interactions. Students are classified into one of the 10 deciles of ability distribution and indicators of corresponding deciles are interacted with the fractions of peers in each of the 10 deciles. This way, Hoxby and Weingarth (2006) can test for substitutability and complementarity between all possible pairs of students and their peers. They reject linear-in-means model and find evidence of monotonicity and benefits of tracking - the largest positive effects are observed from peers on classmates from the same achievement decile.

Burke and Sass (2012) and Imberman et al (2012) use the same strategy to analyse non-linearities and share one common finding - large and significant non-linear effects as compared to standard linear specification, and monotonicity. Both studies reject linear-in-means models as not being supported by the data. Like Hoxby and Weingarth (2006), the study of Imberman et al finds weak evidence of the benefits of tracking. Making use of the available data on school attendance and indicators of students' behavior, Imberman et al (2012) show that disruptive classmates negatively affect discipline and attendance, but not academic performance.

Studies by Lavy, Silva and Weinhardt (2012) and Lavy, Passerman and Schlosser (2012) rely on the variation in the shares of low-achievers and repeaters to estimate effects from "bad" peers. According to their findings, peer effects take the non-linear form with bad peers negatively affecting achievement of everyone but especially of those who are at the lower end of ability distribution. The average effect of bad peers, however, hides the important gender differences - girls do benefit from bright peers while boys are losing out (Lavy, Silva and Weinhardt, 2012).

The overall consensus in the recent literature on peer effects in education is clear - the data do not support the simple linear-in-means model. The evidence suggests that the structure and nature of peer effects in elementary and middle school are more complicated than the standard linear-in-means model implies.

Data

The data in this paper were acquired from the Education Quality and Accountability Office (EQAO), a government agency in the largest Canadian province, Ontario, that designs and oversees provincial standardized tests. The data set comprises a three-year panel of Ontario public schools with individual records for all students in Grade 6. The students' records include their achievement level on standardized provincial tests in Grade 3 and Grade 6 in mathematics, reading and writing and a number of demographic variables The standardized tests, also known as EQAO tests, are designed to evaluate whether a student is prepared to work at the next grade level and how well she or he performs relative to provincial expectations. Unlike many other standardized tests that report results on a scale from 0 to 100, EQAO testing system uses four levels of achievement, from 1 to 4, plus 0. Level 0 indicates that student's achievement falls behind provincial expectations and requires remediation. Not meeting provincial expectations does not prevent students from progressing to the next grade. With the exception of Grade 10 literacy test, students are not required to retake the test if they failed it.

The EQAO tests are low-stake assessments for both students and teachers. The results of these tests do not impact the school grades the child receives and are not recorded in the report card - the complete information of academic performance for each Ontario public school student. As a result, students do not have incentives to exert unusually high efforts, and teachers do not need to teach for the test. That specific nature of the EQAO tests allows using them as a proxy for the ability of a child and all other inputs from parents, teachers and schools accumulated by the time the test was taken. An additional argument in favour of using EQAO test score as a measure of ability is the robust correlation of Grade 3 and Grade 6 test results for individual students. The predictive power of Grade 3 test score is the same for new students and those students who took grade 3 and grade 6 tests in the same school, even though the new students took tests while in different schools.

After imposing a number of sample restrictions, the final data set consists of 228,947 individual student records who represent 12,556 classrooms in an unbalanced panel of 8,135 school-year observations.

Since the identification strategy in this paper exploits the entry of new students to a school, Table 1 describes the academic achievement measured by the EQAO test scores in Grade 3 and 6 and is broken down by incumbent students (those who stayed in the same school from at least Grade 3) and new students (those who entered their current school in the beginning of Grade 6).The first observation from the table is that the incumbent students are doing better than movers on average for all measures of achievement - Grade 3 and 6 test scores and proportion of high achieving students. This should not be surprising given the previous findings in the literature. Thus, Hanushek, Kain and Rivkin (2004) show that student mobility has a negative impact on achievement. The potential reason for lower academic outcomes for movers is that events that make students move – family break-ups, unemployment or loss of parents – also negatively affect their academic achievement (Rothstein 2009, 2010). In addition, it has been shown that student's performance suffers in the first year they move to a new school (Bifulco and Ladd 2006, Sass 2006).

Compared to new students, the distribution of test scores among incumbent students represented by a class standard deviation is much tighter. This is not surprising as they represent a larger fraction of the student body. The proportion of high achievers is larger among incumbent students than among movers, and the share of low achievers is smaller. The distribution of achievement between Grades 3 and 6 changes for both new and incumbent students. Unlike Grade 3 where the majority of students belong to middle-achievers, in Grade 6 more students move to lower or upper tails of the distribution, and there appear to be a smaller number of average students than before. With such divergence from the middle it becomes easier for teachers and principals to stream students by abilities within and between classroom. Unlike in elementary and middle school, tracking by abilities is practiced in Ontario schools.

On average, a classroom in my sample has about 12% of new students - those who entered their current school in Grade 6. With respect to other characteristics available in the data, new students are similar to incumbents. Thus, the shares of girls, foreign-born students and native-speakers are on average the same; the proportion of newcomers, i.e. those six-graders who learn English as a second language, is higher among new students.

Methodology

Identification of peer effects is known to be plagued with issues of selection and reciprocity. In order to overcome those lingering methodological problems, I use instrumental variables strategy. The main assumption underlying this strategy is that the new students to a school are plausibly randomly assigned to classes within a school. If this identifying assumption is satisfied, then the average ability of new students random;y shifts in the average quality of all classmates for an individual student and induces random variation in peer quality across classrooms. It is that variation that I am using to identify the impact of classmates' average quality on individual achievement. In Table 2 I show empirically that students who are new to a school are assigned to classes based on observed characteristics (gender and English as a second language status), but not on ability as measured by the results of the EQAO tests. The first column of Table 2 demonstrates that there is no correlation conditional on observed characteristics of a student, between a new student test score in Grade 3 and average achievement of his or her classmates in a new school. I then run a placebo test when I randomly assign status of a 'new" student to some of the incumbent students. Results of that falsification test presents in column 2 of Table 2. According to the estimates, achievement levels of classmates and individual achievement are negatively related which implies that there is mixing of abilities in a classroom. I confirmed this empirical finding in my personal interviews with school principals and online survey: the overwhelming majority of principals that participated confirmed that they have a policy of mixing students with different level of achievement in classes. However, they also noted that they balance gender composition in a classroom and group students who learn English as a second language. I test and confirmed the last two findings from the survey and interviews in the data: columns 5 and 6 demonstrate that a new girl is more likely to end up in a class with a majority of boys and a new students who learns English as a second language will be assigned to a class with a higher than average share of English as a second language learners. Based on these empirical observations that found support through my interviews with school principals, I use the average ability of new students as a plausibly exogenous shock to the composition of all classmates. The fraction of new students in a class varies across schools and classrooms but is sufficiently large to generate sizable variation in the mean ability of classmates after the entry of new students.

The most commonly used linear model of peer interactions in the literature includes the average of peers' outcomes or characteristics in group j excluding student i outcome and relates that average to the individual outcome of student i:

$$Y_{ij} = \beta_1 \bar{Y}_{(-i)j} + \varepsilon_{ij} \tag{1}$$

Specifically, for student i in classroom c from school s in academic year t the effect of the average quality of classmates measured by their test score at time t the coefficient β_1 represents the linear effect of peers:

$$Y_{icst} = \beta_0 + \beta_1 \bar{Y}_{(-i)cs,t} + \mathbf{X}\gamma + \varepsilon_{\mathbf{icst}}$$

where vector $\mathbf{X}$ represents all available observed characteristics of an individual student as well as characteristics of the classroom, school and neighborhood.

Instrumenting the average test score in class in Grade 6 with the average lagged test score of only new students in the same class, I estimate the effect of peers' ability on individual achievement of incumbent students.

Using this strategy, I find positive, large and statistically significant peer effects at the level of the classroom and no effect at a grade level. Results are presented in Table 3. On average, a one standard deviation increase in peers' quality leads to 0.25 standard deviation improvement in own test score in Grade 6. Analysing results from different specifications - OLS, fixed effects and 2SLS, I find that school administrators mix students into classes based on demographic characteristics and on the achievement level once it is observed. Thus, incumbent students are more likely to be in a class with lower average achievement if they are high-achievers themselves. At the same time, selection into school is positive as indicated by the comparison of OLS and fixed effects models where the latter controls for school and year unobserved heterogeneity and school linear time trends.

In the following sections, I relax the assumptions of the standard linear-in-means model and explore potential non-linearities in peer interactions among students in a classroom.

Effect of different classmates on individual achievement

Since the average effect may hide non-linearities in peer interactions, I relax the linear-in-means specification to disentangle the effect of different peers on own achievement. To do so, I break down the average effect into four different effects resulting from different peers in accordance with four levels of achievement on provincial tests. I use the fractions of classmates in each of the achievement levels to construct four peer variables for each incumbent student.[2] The level of achievement serves as an indicator of an individual student's type. The shares of classmates of each type represent the impact of low, middle or high ability peers on the gain in test score of incumbent students.

I estimate the following equation by the two stage least squares (2SLS) procedure where the fraction of current Grade 6 classmates of each type as indicated by their test score is predicted with the fraction of new students of the same type measure by their achievement in Grade 3:

$$Y_{icst} = \beta_0 + \sum_{j=1}^{4} \beta_j \text{Fraction}^{j}_{(-i)cst} + \mathbf{X}\gamma + \varepsilon_{\mathbf{icst}}$$

[2]The average achievement in the classroom is the sum of the achievement levels weighted by the fraction of students in that level, i.e $\bar{Y} = \sum_{i=1}^{4} \text{Fraction}_i \times \text{Level}_i$.

where $\text{Fraction}^j_{(-i)cst}$ is a share of classmates of student i in class c in school s in academic year t whose achievement level in Grade 6 is j. This empirical specification is similar in fashion to Lavy, Silva and Weinhardt (2012) who estimate a within-pupils regression for three different subjects and differentiate between the impact of peers in the 5% top and bottom percentiles. They find that it's bad peers that matter, not the good ones - the effect of the high ability peers is small and insignificant while the impact of low-achieving classmates is negative and significant and the results are robust to the variety of specifications. Their results are also consistent with the findings in Imberman, Kugler and Sacerdote (2012) for elementary school students.

The first panel of Table 4 shows the effect of different peers on the gain in the math test score for incumbent students. The effect is assumed to be independent of own lagged achievement level. The first and the second columns present OLS and fixed effects estimates respectively. Column 3 shows the reduced form coefficients when peer variable is defined as the fraction of new classmates at a given level of achievement, and column 4 shows the estimates from the 2SLS procedure.

The interpretation of the coefficients is not straightforward. In panel A, I omit the share of Level 2 students due to collinearity. The coefficients then represent the net effect of increasing the share of peers at a given level of achievement while reducing the share of classmates at level 2 and holding constant the two remaining shares. In Panel B, I omit the share of Level 3 peers, so that the coefficients represent the change in the test score as a result of replacing Level 3 classmates with peers from given level of achievement. Finally, Panel C shows the effects when Level 1 and 2 students are combined into one ability group and the net effect is measured relative to Level 3 classmates.

Without any correction for selection into schools and classrooms in column 1 of Table 4, we observe that replacing low-achievers with high-achievers in a classroom would have a large, positive and significant impact on all incumbent students. At the same time, an increase in the share of low-achievers at the expense of average students would negatively affect performance of all students. Fixed effect estimates, while smaller in magnitude, tell the same story. This is because one needs to account for selection into both schools and classrooms when dealing with classroom peers. In that case, school and year fixed effects will not do much for the estimate of the spillovers within the classroom. Moving to the reduced form coefficients in column 3, they confirm that all students in class benefits from good classmates. At the same time, the academic performance of all students suffers from low-ability students. The coefficients, however, are smaller in magnitude, reflecting to some extent the small fractions of new students used to estimate these effects. As opposed to fixed effects and reduced form coefficients, 2SLS estimates present a slightly different picture of ability spillovers in the classroom. The last column of Table 4 implies that once the non-random placement of students into classes within a school is accounted for, the only significant effect we observe is from the high-achieving peers, and not from the low-achieving ones.

An alternative way of thinking about how the composition of peers with different abilities affects own achievement is to represent the average as a weighted sum of achievement levels. For instance, an increase in the average ability of classmates by 0.5 units can be a result of replacing students from the lower end of ability distribution by high-achieving peers. At the same time, it can be the result of the substitution of both very good and very bad students by average students, so that a class is now heavily weighted towards average peers. If the effect from improving average achievement level of classmates is the same as switching of classmates with different abilities required to achieve the new average, then this exercise does not provide any new information about peer effects in addition to the average effect. For instance, if an increase in the share of low achievers and an equivalent decrease in the share of high achievers result in the same reduction in own test score as that from the decrease in the mean ability of classmates, then the observed spillover effect is the result of the overall change of the classmates' quality and has nothing to do with introducing "bad" or "good" peers to a class. In other words, it should not matter whether an improvement in the average ability in a class is caused by the reduction in the share of low-achievers or increase in the fraction of high-achievers.

Table 5 compares the predicted effects from increase or decrease in the average classmates' quality and changes in the composition of the peer group that would bring the equivalent change in the average quality of peers in the classroom. The distribution of the achievement levels in the population of six-graders in Ontario schools is almost perfectly matched with the average observed shares of students in the classroom, so that the classrooms are balanced to represent students from all achievement segments. Table 1 shows that the shares of students at Level 1, 2, 3 and 4 in the entire population of six-graders in Ontario public schools are 6%, 29%, 51% and 14% respectively. Given that distribution, the average math test score in a classroom is 2.73. In Table 5, I set the initial allocation of achievement levels in the classroom according to the distribution above to 5% of Level 1 students, 30% of Level 2 students, 50% of Level 3 and 15% of Level 4 with the average Grade 6 math test score equal to 2.75.

As can be seen from all three panels of the table, the linear-in-means model underestimates the impact of good peers and overestimates the negative effect of the reduction in the average peers' quality on individual achievement gains. For instance, if the average ability in a class goes down, linear-in-means model predicts large and significant negative effect (column 1 of table 5).

Average effect of classmates on different students

In this section, I further relax the standard linear-in-means model of peer effects. The peer variable still represents the average achievement level of classmates, but I also include four interactions of peer variable with indicators of each incumbent student's lagged achievement level. The coefficients on the interaction terms demonstrate the effect of peers' average quality on students from different tails of ability distribution. The estimating equation is:

$$Y_{icst} = \beta_0 + \sum_{j=1}^{4} \beta_{1j} \bar{Y}_{(-i)cs,t} \times D^j_{ics,t-1} + \mathbf{X}\gamma + \varepsilon_{\mathbf{icst}}$$

where Y_{icst} is Grade 6 math test score of incumbent student in class c, school s in the academic year t; $\bar{Y}_{(-i)cs,t}$ is the average achievement of classmates and the dummy variable D^j_{icst} indicates the lagged achievement level of student i. The endogenous terms $\bar{Y}_{(-i)cs,t-1} \times D^j_{ics,t}$ are instrumented with the average achievement of new students interacted with the student i's own level of achievement. This flexible specification also includes a set of cohort, school and school-by-cohort fixed effects, individual and classroom characteristics.

Table 6 reports the achievement level-specific peer effect's coefficients. Panel A shows the reduced form coefficients while panel B presents 2SLS estimates of the average peer ability. All four effects are positive and highly significant with p-values below 0.001. There is variation in the magnitude of the effects for students with different initial achievement. Thus, high-achievers get the highest boost in the test score gain when grouped with high-achieving peers. The lowest gain is for the low-achievers. To put the numbers into perspective, a high-achiever being surrounded by good peers gains a quarter of standard deviation in test score for every standard deviation increase in the average ability of classmates, while a low-achiever gains 0.15 of standard deviation - still a sizable improvement. Results for reading test score (panel C of Table 6) demonstrate the same pattern, with best students gaining the most from good peers, but the gains are of a smaller magnitude. For writing tests (Panel D of Table 6), the low-achieving students seem to gain the most from good classmates, but the rest of the results are similar to mathematics. The difference in the magnitude of the effect size for incumbent students from different levels of ability distribution and the results of χ^2 tests for the equality of the coefficients suggest that there is no reason to believe that linear-in-means model is consistent with the data.

Given that the effect of the average quality of peers is increasing in own ability, it is tempting to conclude that tracking would be an efficient classroom design. However, ability tracking would be beneficial if students of one type benefit from the same type peers more than they do if matched to any other type of classmates. In order to look at the effect of different classmates on different students, in the next section I estimate a two-way interaction model that includes both the shares of peers of different types and the indicators of own ability.

Heterogeneity in response to different peers

The discussion about the benefits of tracking or mixing would be incomplete without taking into consideration the shares of low and high achievers in the classroom in addition to the average achievement and dispersion of ability in the classroom. The move from tracked to mixed classroom is accompanied by a change in the mean test score of classmates, standard deviation of test scores and a shift in the proportions of peers of different types.

In this section, I extend the model to include interactions not only with the average ability of classmates but also with the fraction of new students at each of the four ability levels. This model allows for comparison of the strength of the effects of different classmates, i.e. "good" versus "bad", on incumbent students from different achievement levels. Using the coefficients from the model, I can also look at the complementarity and substitutability between peers and their implications for ability tracking in schools.

I estimate the following equation by 2SLS procedure, where the fraction of current Grade 6 classmate at each level of achievement is predicted with the fraction of only new students at the same level of achievement.

$$Y_{icst} = \beta_0 + \sum_{k=1,3,4} \sum_{j=1}^{4} \beta_{jk} \text{Fraction}^j_{(-i)cst} \times D^j_{ics,t-1} + \mathbf{X}\gamma + \varepsilon_{\mathbf{icst}}$$

where $\text{Fraction}^j_{(-i)cst}$ is a share of classmates of student i in class c in school s in academic year t whose achievement level in Grade 6 is j. This specification of the peer effects model accounts for all possible interactions between an incumbent student ability and shares of classmates at a given level of achievement.

The intuition and assumptions behind the identification strategy are the same as before. If new students are placed into classes not based on their lagged test score, then this allocation generates a plausibly random change in the composition of current classmates. For instance, a class with an initially even distribution of students' achievement (25% of each) gets three students; two of them are high achievers and one is a low achiever. Then, the shares of students in that classroom would be altered - the fractions of high and low achievers will go up for a plausibly exogenous reason and shares of the average students will go down. Such change in the composition of classmates might have a different effect on students with different abilities through various channels. For instance, a teacher might need to adjust her instruction to tailor it to the largest share of students in class - the high achievers. That might have an adverse impact on low achievers and even on the average students. At the same time, if there are spillovers from good students, then a larger share of high-achievers would have a positive impact on everyone in the classroom.

Table 7 reports results of the estimation of the two-way interactions model. The two panels of the table present results for the mathematics tests score from the reduced form and 2SLS respectively.

The first column of table 7 reproduces results from Table 2 for comparison. Recall from the previous sections that all students benefit from good peers, and this effect is statistically significant and remains large, independent of specification. Holding the fraction of classmates at level 1 and 3 constant and substituting away peers at level 2 results in math test score gains for all students except level 2 students. The IV estimate is larger in magnitude than its reduced form counterpart. The IV estimates also imply that changing the share of low and middle achievers has no effect on test score gain of incumbent students (while large in magnitude, the coefficients are not statistically significant). The corresponding reduced form estimates are both significant, but smaller in magnitude.

Moving to the effect of different peers on different types of incumbent students, Table 7 reports 12 coefficients of the reduced form and 2SLS estimates. Only four of the 2SLS estimates are significant and almost all of them represent the effects of increasing the share of high-achieving peers which are consistent with the aggregate effects in the first column. For instance, increase in the share of high-achievers by 10% and reduction in the fraction of classmates at level 2 will raises the achievement of high-ability students by 0.084 points. This statistically significant result seems to favour ability tracking as the efficient strategy of classroom organization. However, low achievers benefit four times more from the same increase in the share of high-achieving classmates - the improvement in test score for them is 0.351 points. The low-achievers do not benefit nor do they suffer from the increase in the share of low-achieving peers or those in the middle of the distribution.

Low achievers gain when "average" classmates are substituted by either "good" or "bad" peers.[4] It is not surprising to find that "bad" students benefit from a larger proportion of high achieving peers. For instance, good students might serve as a role model for the low achievers, or teachers might increase their expectations when the fraction of high achievers is larger, and this would motivate the "bad" students to learn and perform better. However, the opposite also might be true if low achievers are marginalized in a classroom. The fact that low achievers are better off in the presence of similar classmates is also not unusual. It might be that in a more homogeneous class no one is "left behind" and everyone acquires knowledge at the same pace. The response of high achievers (column 5) is very different. I find that "good" students are "immune" to the shifts in the composition of the classmates. It is reasonable to assume that high achieving students are also independent learners and their performance does not depend on the composition of the classroom, but only on the average achievement level of the classmates. "Marginal" students benefit the most when the fraction of students in the middle of the distribution goes up, so they are better off when grouped with similar students. "Average" students show better results when the fraction of high achieving peers increases.

Marginal students perform worse if they are grouped with low-achieving peers. The effect is large and significant. Marginal students, unlike others, do not benefit from top students and seem to prefer being grouped with similar students. Average achievers do better when surrounded by excellent peers and experience no change in test score when middle-achieving student is substituted by someone from the low end of ability distribution. Same holds for high-achievers - while they gain from replacing a low or/and middle achiever by a top student, they are indifferent to changes among students who are lower achievers than themselves.

One important implication of the analysis above is that the progress of the majority of the students is not hampered by the presence of low-achieving classmates, and almost everyone is doing better surrounded by top students, with low-achievers benefiting the most.

For convenience, I reproduce these results in a matrix replacing insignificant coefficient with zeroes, positive and significant with a plus sign, and negative and significant with a minus sign. This representation would allow to clearly see the patterns of peer effects when discussing the models of peer interactions.

	Level 1	Level 2	Level 3	Level 4
Share level 1	0	−	0	0
Share level 3	0	0	0	0
Share level 4	+	0	+	+

To make the interpretation of the marginal effects more clear-cut, I also estimated a model where the fraction of peers in the middle of the achievement distribution is held constant and the shares of low and high ability students are allowed to vary. I combined shares of students at level 1 and 2 and omitted the share of students at level 3. The results are presented in Table 8 and in a matrix of effects below.

	Level 1	Level 2	Level 3	Level 4
Share level 1&2	+	−	−	0
Share level 4	+	0	+	+

The results are in line with Table 7: everyone gains when a low-achieving student is replaced by a high-achiever, but not everyone loses when a high-achiever is replaced by a low-achiever. For instance, students who are low-achievers themselves, gain when the share of low ability students increases and the share of high-achievers decreases. Increase in the share of low-achievers has no impact on top students and has a negative effect on students in the middle. As shown before, students who scored just below the provincial standards prefer to stay with the classmates who are at the same achievement level. One explanation is that increasing the share of either good or bad students refocuses teacher's attention to those students, leaving out the students from the middle of the distribution. Another way to think about it is the analogy with the median voter: once the class is balanced, the instruction is targeted at the median student, but once the class composition changes, the teacher needs to adjust the pace and difficulty of the material to target the "new" median student.

Policy Experiment

The coefficient estimates in the heterogeneous peer effect model allow me to simulate a policy experiment where I vary the fraction of students in a class and observe the predicted impact on individual student's test score. I assume that the initial allocation of students is consistent with the observed distribution of abilities in the population of Ontario 3rd graders: 20 percent level 1 and 2 students, 60 percent level 3 students and 20 percent level 4 students. Table 9 reports the estimated effect of moving from that allocation to tracking classes with high proportions of low or high achievers. This immediate effect does not include the impact on students who are replaced away from the classes, and does not represent the general equilibrium effects. The experiment also assumes the constant distribution of abilities in the population.

In panel A, the fraction of low achievers goes up and shares of both middle and high achievers go down. This change represents a shift to tracked classroom where the majority of students (60%) are from the lower end of ability distribution. As can be seen from the estimates, the gain for the low achievers from such allocation cannot offset the losses for other students in the same classroom.

For the next experiment, Panel B, the class is dominated by high-achievers. Everyone independent of their lagged test score benefits from being in class with high-achievers, with the low ability students gaining the most.

In the last panel of Table 9 the class is heavily weighted towards the middle ability students. Removing students from the low and high ends of ability distribution is beneficial only for marginal students (those who were at Level 2 of provincial achievement in Grade 3). All other students experience losses in their test score. Again, the net losses or gains cannot be estimated from that type of experiment, as these do not include the effect on students who left the classroom.

Diversity of abilities in the classroom

The discussion about the differential effect of different classmates on individual achievement would be incomplete without mentioning the effect of the overall diversity of abilities in the classroom. This section presents the estimates of the effect of class heterogeneity measured by the variance in classmates' test scores. The estimating equation is again a standard peer effects model that assumes linear relationship between the average class characteristic and individual achievement gain.

$$Y_{icst} = \beta_0 + \beta_1 \bar{W}_{(-i)cs,t} + \mathbf{X}\gamma + \varepsilon_{\mathbf{icst}}$$

where dependent variable is Grade 6 test score of student i in classroom c in school s during the academic year t. $\bar{W}_{(-i)cs,t}$ represents the classroom heterogeneity and is measured by the standard deviation of test scores for all students in a classroom except student i contribution. X is a vector of controls including lagged test score of student i, gender, special education status, whether student is in French immersion or English as a second language program.

Regression also included school, year and school-by-year fixed effects to control for selection into schools. An error term, ε_{icst}, represents the aggregated unobserved heterogeneity at school, class and individual levels. Grade 6 class standard deviation $\bar{W}_{(-i)cst}$ which is an endogenous term, is predicted with the standard deviation of new students' Grade 3 test score in the same class.

As seen in the previously estimated models with shares of different peers, the compositional effect of classmates is different from the average quality effect. The regressions with shares are informative to assess the impact of bad or good peers on individual achievement, but they are silent about the effect of diversity in a classroom on academic performance. In general, class diversity might affect individual achievement through class disruption or tailored teacher instruction, among other channels. Table 10 demonstrates the impact of class heterogeneity on individual test score. According to fixed effects regression in column (1) of Table 10, class heterogeneity has a positive and significant effect on the gain in test score from Grade 3 to Grade 6 independent of whether regression also includes average peer ability. This magnitude of the effect is large relative to that of the average peer ability. However, the coefficient estimates are likely to be biased because school and year fixed effects only capture selection of students into schools and control for between school heterogeneity. As has been mentioned before, 2SLS estimates show that there is negative sorting of students into classes within a school based on observed lagged ability. Moreover, it follows from both the quantitative evidence and personal interviews with school principals that students are grouped into classes in a way to create an inclusive environment and heterogeneous classrooms. Students in Ontario public schools seemed to be allocated to classrooms based on a variety of characteristics but using the principle of complementarity.

Since class dispersion of test scores reflects the endogenous decision of school principals and parents to place a student into a mixed or tracked classroom (Hanushek *et al* 2003), the coefficient in the fixed effects model should be interpreted with caution. In order to learn about the bias in the fixed effect estimate, I instrument class standard deviation with the standard deviation of lagged test score for new classmates only (column (2) of Table 10). While the effect of the average peer quality retains its significance and even increases in magnitude, the effect of class heterogeneity vanishes completely. Similar pattern arises in the reduced form estimation where dispersion among new students enters the equation directly (column (3) of Table 10). Previous studies that use standard deviation as a measure of heterogeneity find negative or no impact on test scores (Burke and Sass 2013 and Duflo *et al* 2008) except Vigdor and Nechyba (2005) who have estimated positive effect of ability dispersion on test scores among 5th graders in North Carolina schools. The conventional explanation for the negative effect is that it is difficult to teach effectively and tailor instruction level to students with disperse abilities in the subject. Thus, the finding that class heterogeneity hinders the achievement gains is suggestive of benefits of tracking, i.e. that students with similar abilities benefit from learning in the same environment. The finding of a positive effect, on the contrary, suggests that mixing students with different ability levels in the same classroom might be beneficial for all students independent of own lagged achievement. The positive effect of diversity of abilities in the classroom supports the idea that students learn from each other and not only from their teacher, and class interactions play an important role in learning and transmission of knowledge.

To test further for the effect of class heterogeneity on students with different initial ability, Table 10 lists coefficients on interactions of own ability and class standard deviation. The last four columns of Table 8 are more informative, as they show the impact of dispersion on different students. For instance, high achievers (those who achieved level 4 at Grade 3 math test) do benefit from the overall class heterogeneity, while all other students are not affected by that characteristic at all. This pattern holds when the average achievement of classmates is included in the regression.

The absence of the effect of class heterogeneity on individual test score gain implies that both tracking by abilities or mixing students with different abilities in the same class are not a uniform solution for raising individual achievement. However, taken together, the results from this and previous section make it clear that all students benefit from the presence of high-achieving classmates. In the next section, I analyze how the results from the models that assume heterogeneous effect of classmates are aligned with the theoretical models of peer effects.

Models of peer effects

The coefficients from the fully saturated model and the model with class dispersion, their signs and magnitude inform about the consistency of the data with the models of peer effects. Hoxby and Weingarth (2006) provided the most comprehensive description of peer effects' models which they test using the data from the Wake County school district. Burke and Sass (2013) and Imberman *et al* (2012) test for the presence of heterogeneous peer effects, and Sacerdote (2011) summarizes the recent findings in the education peer effects literature. I briefly describe models of peer interactions and show which of these models are supported by the data in present research.[5]

The linear-in-means model represents the simplest relationship between peers and own ability. The effect of the average ability is assumed to be independent of the own ability or achievement level of a student. That model is unambiguously rejected, as demonstrated by the results and tests in Table 4. The other two models that also assume homogeneous effect of peers on all students are the shining light model and the bad apple model. In the shining light model, one bright student serves as a powerful role model for others in the same class, and other students try to mimic his or her behavior. In the empirical estimation that would be the case when high ability peers have the same effect on everyone regardless of own achievement of a student. This is clearly the case among Ontario six-graders. The positive effect of good peers on everyone is the robust finding across all specifications. The bad apple model is the shining light with the minus sign, i.e. one disruptive student in a class is enough to harm the achievement of every child. For that model to hold, the data must indicate that low ability students have the same negative effect on test score gains for everyone. The bad apple model is not supported by the data since increasing the share of low-achievers in the classroom does not affect high-achievers at all, and even has a positive impact on low-achievers themselves.

A number of peer interactions' models are based on the impact of the standard deviation of peer characteristics, and not on the average quality. The two common models of class dispersion are focus model and rainbow model. The focus model is based on assumption that the smaller variability in achievement levels in the classroom is beneficial for the academic progress. The rainbow model, as opposed to the focus model, implies that diversity in class generates better outcomes for every student. The standard test of these models involves inclusion of the class variance into the regression relating an individual test score and average achievement of the classmates. If the coefficient on classroom dispersion of test scores is negative, it implies that homogeneity is good for everyone no matter the ability of individual student, and that the data support the focus model of peer effects. If the coefficient is positive, i.e. classroom heterogeneity is beneficial for every student in class, this finding would favor the rainbow model. Both models assume the same effect on everyone independent of the initial ability or achievement. Given the results of in Table 10, after correcting for selection into schools and classrooms, I do not find significant impact of class dispersion on gains in math test score among Ontario six-graders.

The next set of models assumes heterogeneous effect of peers on students with different initial ability. The first one – a boutique model – implies that students benefit from peers with the same level of ability. For instance, being surrounded by low ability classmates is beneficial for low-achievers while being harmful for everyone else. The same should be true for high and middle ability students. If this model finds support in the data together with the focus model, the obvious implication would be that tracking by abilities is the most efficient design of the classroom. Ability grouping or streaming is regarded to be a useful class organization as teachers may adjust their strategies to cater to the uniform group of students and raise or decrease the expectations target, but it is not a common practice in elementary schools. Given the empirical specification and results in Table 4, the test for the tracking model of peer interactions would be rejected if the increase in the share of peers from the same level of achievement as student's own has smaller impact than the increase in the share of higher achieving peers. For instance, if having more peers from Level 3 than from Level 4 is more beneficial for Level 3 students, and having more peers from Level 1 than from Levels 2, 3, and 4 is better for Level 1 students, then the tracking or boutique model is supported by the data. While grouping by ability seems to be beneficial for high-achieving students (Level 4), there is no evidence to support the tracking model for all other students. Previous studies by Hoxby and Weingarth (2006), Imberman et al. (2012) and Burke and Sass (2013) also found little support for the boutique model.

The next model of heterogeneous peer effects – the invidious comparison model – borrows intuition from the behavioral and sociological literature and suggests that students are performing worse in the presence of higher achieving peers. In other words, an increase in the share of peers whose Grade 3 test score was higher than the student's own score leads to the reduction in Grade 6 test score for a given student. And at the same time, when the share of low achieving peers goes up, then that student's academic performance improves. Similar to tracking, invidious comparison model is not supported by the data in this study. All students independent of their own ability seem to benefit from an increase in the proportion of high ability peers and suffer from an increase in the share of low achievers, except for the low achieving students themselves. The impact from an increase in the share of high achieving peers (Level 4) is positive and significant for both low-achievers and those who are in the middle of the ability distribution.

The model that most often finds support in the data is a single-crossing model: the positive effect of high-ability classmates is increasing in own ability of a student. This model is also known as monotonicity property of peer effects. I do find that Ontario data is consistent with the monotonicity property when I allow the effect of peers to differ for students with different abilities. When I estimate the model with shares, I find that low-achievers gain more than high achievers when a low ability student is replaced by a top student in the same class. To reconcile these findings, I hypothesise that while the average quality of peers is more important for high-achievers, adding just one more smart kid in a classroom has a larger impact on marginal kids than it has on top students.

The overall findings imply that the structure of peer effects in elementary school is more complex than suggested by the simple linear-in-means model. While all students independent of their own ability benefit from the presence of high-achieving peers, this effect is different in magnitude. I find little evidence that tracking by ability would benefit elementary students. The only model that finds support in the data is a single-crossing model in its weakest version, monotonicity.

Conclusion

In this paper, I analyzed non-linearities in peer interactions among Ontario public school students. I exploited the entry of new students to a school as a plausibly exogenous shift in the composition of the classmates.

Besides positive and significant average effect described in the first chapter, I find that the peer group composition matters. For instance, being surrounded by good peers is beneficial for everyone independent of their own achievement level. The presence of low-achievers in a class does not impede the achievement of other students, and even helps students who are low achievers themselves. The overall diversity of the classroom by itself does not seem to be a factor in determining academic progress.

The results of the fully saturated model of peer interactions are consistent with the monotonicity model of peer effects, where the average effect from good peers is increasing in own ability. This observation may point out that tracking is an efficient model of organizing students into classes within a school. However, removing good peers from an average class would hurt low-achievers more than it would help high achievers, as evidenced by a hypothetical policy experiment.

Overall, the findings in this chapter favour the idea that peer effects in schools are more complicated than the linear-in-means model implies, and policy interventions should take into consideration the diversity of the response of different students to different peers.

Notes

[1]For a review of literature on tracking and streaming, see J.Betts (2011).

[2]See, for instance, works of J. Oakes for discussion about tracking and inequality in the US education.

[3]School efficiency resulting from ability tracking is measured by the differences in the average achievement in tracked and non-tracked schools. Inequality is the difference in the outcomes for students with different initial ability.

[4]In order to make the interpretation easier, instead of labeling students by the level of achievement as 1 to 4, I will call students at the lowest level of achievement "bad" students without attaching the actual meaning of the word "bad"; students at the highest level – "good" students, and students in the middle of the achievement distribution - "average". Among "average" students, I will distinguish between "marginal" (those whose achievement is below provincial standards, or level 2) and just "average" (level 3).

[5]This section is based on description and classification of peer effects models from Hoxby and Weingarth (2006).

References

[1] Ammermller, A. (2005). Educational opportunities and the role of institutions. Centre for European Economic Research Discussion Paper No.05-44.

[2] Argys, Laura M., Rees D. and Brewer, D.(1996). Detracking America's schools: equity at zero cost?. *Journal of Policy Analysis and Management*, 15, no. 4,623-645.

[3] Betts, J. (2011). The Economics of tracking in education. Handbook of Economics of Education,Volume 3, Hanushek, E., S. Machin and L. Woößmann, eds., Amsterdam: North Holland, 341-381.

[4] Betts, J. and Shkolnik,J. (2000b). Key difficulties in identifying the effeects of ability grouping on student achievement.*Economics of Education Review*, 19, no.1, 21-26.

[5] Bifulco, R. and Ladd, H. (2006). Institutional Cahnge and Co-Production of Public Services: The Effect of charter School on Parental Involvement. *Journal of Public Administration Reserach and Theory*, 16, 553–576.

[6] Brunello, G. and Checchi, D. (2007). Does school tracking affect equality of opportunity? New international evidence. *Economic Policy*, 52, 781–861.

[7] Burke, M.A. and Sass, T.R. (2013). Classroom peer effects and student achievement. *Journal of Labor Economics*, 31(1), 51-82.

[8] Duflo, E., Dupas, P. and Kremer, M. (2011). Peer Effects, Teacher Incentives, and the Impact of Tracking: Evidence from a Randomized Evaluation in Kenya. *American Economic Review*, 101(5), 1739–74.

[9] Epple, D., and Romano, R. (1998). Competition between private and public schools, vouchers, and peer-group effects. *American Economic Review*, 88, 33-62.

[10] Epple, D., Newlon, E. and Romano, D. (2002). Ability tracking, school competition and the distribution of educational benefits. *Journal of Public Economics*, 83, 1-48.

[11] Gamoran,A., and Mare, R. (1989). Secondary school tracking and educational inequality: compensation, reinforcement, or neutrality? *American Journal of Sociology*, 94, no.5, 1146-1183.

[12] Hanushek, E.A. and Woößmann, L. (2012). Do better schools lead to more growth? Cognitive skills, economic outcomes, and causation. *Journal of Economic Growth*, vol. 17(4), pages 267–321.

[13] Hoffer, T. (1992). Middle school ability grouping and student achievement in science and mathematics. *Educational Evaluation and Policy Analysis*, 14, no.3, 205–227.

[14] Hoxby, C., Weingarth G. (2006). Taking race out of the equation: School reassignment and the structure of peer effects. Harvard University, unpublished manuscript.

[15] Imberman, S. A., Kugler, A. & Sacerdote, B. (2012). Katrina's children: Evidence on the structure of peer effects from hurricane evacuees. *American Economic Review*, vol. 102 (5),2048-82.

[16] Lavy, V., Passerman, D. and Schlosser, A. (2012). Inside the black box of ability peer effects: Evidence from variation in low achievers in the classroom. *Economic Journal*, 122, 208–237.

[17] Lavy, V., Silva, O. & Weinhardt, F. (2012). The good, the bad and the average: Evidence of ability peer effects in schools. *Journal of Labor Economics*, 30 (2), 367-414.

[18] Meghir,C. and Palme, M. (2005). Educational reform, ability, and family background.*American Economic Review*, 95(1), 414-424.

[19] Oakes,J. (2005). Keeping track: how schools structure inequality. New Haven: Yale Univeristy Press.

[20] Pekkarinen, T., Uusitalo, R. and Pekkala, S. (2006). Education policy and intergenerational income mobility: evidence from the Finnish comprehensive shool reform. IZA Discussion Paper No. 2204.

[21] Rothstein, J. (2006). Good principals or good peers: Parental valuation of school characteristics, Tiebout equilibrium, and the incentive effects of competition among jurisdictions. *American Economic Review* 96(4), 1333-1350.

[22] Rothstein, J. (2009). Student Sorting and Bias in Value Added Estimation: Selection on Observables and Unobservables. *Education Finance and Policy* 4(4), 537–571.

[23] rothstein2010 Rothstein, J. (2010). Teacher Quality in Educational Production: Tracking, Decay, and Student Achievement. *Quarterly Journal of Economics* 125(1), 175–214.

[24] Sacerdote, B. (2011). Peer effects in education: how might they work, how big they are and how much we know thus far? in *Handbook of Economics of Education*, volume 3.

[25] Sass, T. (2006). Charter Schools and Student Achievement in Florida. *Education Finance and Policy*, 91-122.

[26] Vigdor, J.L. and Nechyba, T.S. (2007). Peer effects in North Carolina public schools. In: Woessmann, L., Peterson, P.E. (*Eds.*), Schools and the Equal Opportunity Problem. The MIT Press, 73–101.

Table 1. Summary Statistics (mathematics test scores)

	Incumbent students		New students		All Students	
	Grade 3	Grade 6	Grade 3	Grade 6	Grade 3	Grade 6
Average math test score, by classroom	2.796 (0.308)	2.756 (0.395)	2.682 (0.572)	2.569 (0.604)	2.780 (0.296)	2.733 (0.396)
Standard deviation of math test score, by classroom	0.598 (0.154)	0.655 (0.148)	0.530 (0.412)	0.559 (0.410)	0.612 (0.142)	0.663 (0.137)
Fraction of students, Level 1	2.79	5.07	4.51	8.9	3.01	5.56
Fraction of students, Level 2	25.42	28.45	31.97	35.44	26.26	29.34
Fraction of students, Level 3	60.82	52.12	54.96	45.21	60.07	51.24
Fraction of students, Level 4	10.98	14.36	8.56	10.45	10.67	13.86
Number of students	199,717		29,230		228,947	

Note: The sample comprises all students in Grade 6 for three academic years: 2008-2010. The sample is broken down by new and incumbent students. The new student is defined as someone who moved to a new school at the beginning of Grade 6. The averages in cells are the classroom mean and standard deviations for new and incumbent students in the same classroom.

Table 2: Assignment of new students into classrooms

Dependent variable>>>>	Lagged test score of a new student	Lagged test score of incumbent	New student (=1)	New student (=1)	Gender of a new student (female=1)	New student is in ESL program	New student is foreign born	New student did not learn English at home
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Average lagged math test score of incumbent students in class	0.054	-0.205***	0.000	-0.007	0.008	-0.000	0.026	-0.011
	(0.052)	(0.053)	(0.002)	(0.006)	(0.031)	(0.001)	(0.018)	(0.019)
Fraction of female incumbent students in class	0.002				-0.088**			
	(0.067)				(0.041)			
Fraction of ESL incumbent students in class	-2.025					0.604***		
	(3.782)					(0.106)		
Fraction of foreign-born incumbent students in class	0.132						0.004	
	(0.109)						(0.044)	
Fraction of students who did not learn English at home	0.247***							-0.001
	(0.099)							(0.045)
Female *Average lagged math test score of incumbent students	-0.007							
	(0.051)							
ESL *Average lagged math test score of incumbent students	-0.331							
	(1.072)							
Forein born *Average lagged math test score of incumbent sts	0.077							
	(0.086)							
Second language*Average lagged math test score of incumbent sts	0.056							
	(0.075)							
Number of observations	29, 230	29,230	228,947	228,947	29,230	29,230	29,230	29,930

Note: Standard errors clustered at school level. *** p-value<0.01, ** p-value <0.05, * p-value<0.10. Sample includes all students in Grade 6 in classes with a mix of new and incumbent students (columns 3 and 4). In columns (1), (5)-(8) sample includes only new students, and in column (2) sample includes randomly drawn incumbent students. Column (1) reports coefficient estimates from the regression of lagged test score of a new student on the average lagged achievement of all incumbent students in that class. In Column (2) I randomly draw a sample of incumbent students , designate them as new and estimate the same regression as in column (1). In columns (3)-(4) I estimate a linear probabilty model to predict a new student in class. In column (3) new student is defined as someone who entered in Grade 6, in column (4) as someone who entered at any time in grades 4 to 6. Columns (5)-(8) report the coefficients from regression where dependent variable is a characteristic of a new student - gender, ESL (English as a Second Language program), foreign born and did not learn English at home. All regressions include individual and neighbourhood controls. All regressions also include school, cohort and school-by-cohort fixed effects.

Table 3. Effect of the Average Quality of Peers Entered in Grade 6 on Test Scores (IV)

Dependent variable>>>>	Grade 6 mathematics test score of incumbent student				2SLS estimates of the effect of the average Grade 6 classmates test score if the initial level of achivement of incumbent student is:			
	OLS	Reduced Form	2SLS First Stage	2SLS	Level 1	Level 2	Level 3	Level 4
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Average Grade 6 mathematics test score of classmates	0.059***		0.141***	0.419***	0.303***	0.316***	0.406***	0.455***
	(0.013)		(0.002)	(0.053)	(0.074)	(0.059)	(0.054)	(0.055)
Average Grade 3 mathematics test score of new students only		0.055***						
		(0.007)						
Individual controls	Yes	Yes		Yes	Yes	Yes	Yes	Yes
Year fixed effects	Yes	Yes		Yes	Yes	Yes	Yes	Yes
School fixed effects	Yes	Yes		Yes	Yes	Yes	Yes	Yes
School-year fixed effects	Yes	Yes		Yes	Yes	Yes	Yes	Yes
Chi2 test of joint equality of coefficients					220.53***			
Number of observations	199,717	199,717	199,717	199,717	199,717			
R^2	0.30	0.28	0.15	0.36	0.36			

Note: Standard errors clustered at school level. *** p-value<0.01, ** p-value <0.05, * p-value<0.10. Sample consists of incumbent students only, i.e. those who stayed in the same school from at least Grade 3 to Grade 6. The intrument is the average achievement of new students who entered at the start of Grade 6. Regressions cover years 2008-2010 and include individual controls (own test score in Grade 3, gender, Enlglish as Second Language learner, Canadian born, and whether student learned English at home), average score of "old" peers, school controls - urban school, Catholic school board, school from Toronto Metropolitan Area; neighborhood controls - log of median household income, proportion of residents with university degree, proportion of low income families, and proportion of recent immigrants. Column (1) shows OLS estimates when average Grade 6 test score of classmates without student *i* contribution is entered directly into regression. Column (2) shows the estimate of the reduced form specification with average lagged test score of new peers. Columns (3) and (4) present first and second stage estimates of 2SLS where average Grade 6 test score of all students in class without student *i* contribution is instrumented with average lagged achievement of new peers in that class. Columns (5)-(8) shows estimate of linear-in-means model when average test score of classmates is interacted with own lagged achievement level and instrumented with the lagged average test score of new peers interacted with own lagged achievement.

Table 4. Effect of the fractions of students with different achievement levels on gain in test score of incumbent students

	Effect on Grade 6 math test score of incumbent student			
	OLS	OLS	Reduced form	IV
Panel A				
Fraction of students who achieved Level 1 in Grade 3	-0.666***	-0.389***	-0.026**	-0.534
	(0.021)	(0.025)	(0.011)	(0.570)
Fraction of students who achieved Level 3 in Grade 3	0.814***	0.582***	0.025***	0.217
	(0.010)	(0.013)	(0.004)	(0.284)
Fraction of students who achieved Level 4 in Grade 3	1.633***	1.113***	0.049***	0.568***
	(0.011)	(0.016)	(0.008)	(0.184)
Panel B				
Fraction of students who achieved Level 1 in Grade 3	-1.458***	-0.311***	-0.100***	-0.602
	(0.032)	(0.058)	(0.029)	(0.769)
Fraction of students who achieved Level 2 in Grade 3	-0.798***	-0.064*	-0.043***	-0.183
	(0.018)	(0.033)	(0.011)	(0.611)
Fraction of students who achieved Level 4 in Grade 3	0.802***	0.086**	0.091***	0.706**
	(0.021)	(0.039)	(0.017)	(0.278)
Panel C				
Fraction of students who achieved Levels 1&2 in Grade 3	-0.975***	-0.947***	-0.049***	-0.370
	0.015)	(0.029)	(0.011)	(0.296)
Fraction of students who achieved Level 4 in Grade 3	0.784***	0.633***	0.091***	0.721***
	(0.021)	(0.039)	(0.017)	(0.280)
School FE	N	Y	Y	Y
Year FE	N	Y	Y	Y
School-by-year FE	N	Y	Y	Y
Number of observations			199,717	

Note: Standard errors clustered at school level. *** p-value<0.01, ** p-value <0.05, * p-value<0.10. The sample comprises all incumbent students in Grade 3 for three years: 2008-2010. All regressions include school, year and school-by-year fixed effects. The coefficients in the first colum represent the effect of entry of new students with corresponding level of achievement on all incumbent students independent of their achievement level in Grade 3. Reduced form estimates are presented in Panel A when fraction of new students at different levels of achievement enters directly into the estimating equation. Panel B presents the coefficients of the 2SLS procedure when fraction of all classmates in Grade 6 at a specific level of achievement is instrumented with the fraction of new students only at that same level of achievement.

Table 5. Predicted effect of the change in the compostion and average achievement of classmates on incumbent student math test score

	Panel A		
	Linear-in-means model	Fraction of Level 3 peers goes down and Level 1 peers goes up by 12.5%	Fraction of Level 3 peers goes down and Level 2 peers goes up by 25%
Effect of a decrease in the average classmates math test score by 0.25 units from 2.75 to 2.5 on individual test score	-0.105*	-0.075	-0.045
	Panel B		
	Linear-in-means model	Fraction of Level 3 peers goes up and Level 2 peers goes down by 25%	Fraction of Level 3 peers goes down and Level 4 peers goes up by 25%
Effect of an increase in the average classmates math test score by 0.25 units from 2.75 to 3.0 on individual test score	0.105*	0.045	0.177*
	Panel C		
	Linear-in-means model	Fraction of Level 3 peers goes down and Level 4 peers goes up by 50%	
Effect of an increase in the average classmates math test score by 0.5 units from 2.75 to 3.25 on individual test score	0.209*	0.353*	

Note: The numbers in cells show the predicted effects from the changes in the average classmates score as estimated by the linear-in-means model (column 1), or by the change in the shares of classmated from different achievement levels that would result in the same change in the average test score (columns 2 and 3). For linear-in-means model the effect on individual test score gain is computed using the 2SLS coefficient from Table 1a in Appendix multiplied by the change in the average classmates test score. For instance, if the average classmates achievement goes down by 0.5 units, the corresponding change in the individual test score is predicted to be 0.5*0.419=0.209 units. Columns 2 and 3 decompose the 0.25 unit change in the average classmates score into changes in the shares of classmates from different achievement levels. The effects in Columns 2 and 3 and Column 2 in Panel C are computed by multiplying the percentage change in the share of classmates from a given level required for an increase/decrease in the average class score by the 2SLS coefficient estimate from Table 2. For instance, the effect from the reduction in the share of Level 3 classmates by 12.5% with corresponding increase in the share of Level 1 peers by 12.5% is equal to 0.125*(-0.602)=0.075 units. Significant effects are denoted with *.

Table 6. Effect of the average classmates ability on the test score gain of incumbent students

	Effect of the average classmates test score if the initial level of achivement of incumbent student is:			
	Level 1	Level 2	Level 3	Level 4
Panel A (Mathematics -Reduced form)				
Average Grade 3 mathematics test score of new students only × own achievement level	0.060***	0.021**	0.065***	0.074***
	(0.015)	(0.009)	(0.008)	(0.010)
Panel B (Mathematics - IV)				
Average Grade 6 mathematics test score of classmates × own achievement level	0.303***	0.316***	0.406***	0.455***
	(0.074)	(0.059)	(0.054)	(0.055)
Chi-2 test of the joint equality of coefficients	220.53***			
Panel C (Reading - IV)				
Average Grade 6 reading test score of classmates × own achievement level	0.196***	0.250***	0.330***	0.360***
	(0.072)	(0.068)	(0.068)	(0.069)
Chi-2 test of the joint equality of coefficients	192.15***			
Panel D (Writing - IV)				
Average Grade 6 writing test score of classmates × own achievement level	0.438***	0.338***	0.382***	0.398***
	(0.138)	(0.127)	(0.126)	(0.129)
Chi-2 test of the joint equality of coefficients	309.04***			
Individual controls	Yes	Yes	Yes	Yes
Year fixed effects	Yes	Yes	Yes	Yes
School fixed effects	Yes	Yes	Yes	Yes
School-year fixed effects	Yes	Yes	Yes	Yes
Number of observations	199,717			

Note: Standard errors clustered at school level. *** p-value<0.01, ** p-value <0.05, * p-value<0.10. Sample consists of incumbent students only, i.e. those who stayed in the same school from at least Grade 3 to Grade 6. The instrument is the average achievement of new students who entered at the start of Grade 6. Regressions cover years 2008-2010 and include individual controls (own test score in Grade 3, gender, Enlglish as Second Language learner, Canadian born, and whether student learned English at home), average score of "old" peers, school controls - urban school, Catholic school board, school from Toronto Metropolitan Area; neighborhood controls - log of median household income, proportion of residents with university degree, proportion of low income families, and proportion of recent immigrants. Panel A shows the reduced form estimates for math test score gain when the average achievement of new peers is interacted with individula achievement dummy for each incumbent student. Panel B presents 2SLS estimates for the math test score gain when the interactions of the average achievement of all classmates and own achievement are instumented with the interactions of new peers average achievement and own test score in Grade 6. Panels C and D report the coefficients for reading and writing test score respectively.

Table 7. Effect of the classmates on gain in test score for incumbent students with different achievement levels

	All incumbent students	Grade 3 achievement level of incumbent student			
		Level 1	Level 2	Level 3	Level 4
Panel A (Reduced form)					
Fraction of new students who achieved Level 1 in Grade 3	-0.026** (0.011)	0.240*** (0.041)	-0.102*** (0.017)	-0.011 (0.013)	0.006 (0.031)
Fraction of new students who achieved Level 3 in Grade 3	0.025*** (0.004)	0.160*** (0.016)	-0.025*** (0.007)	0.042*** (0.005)	0.015 (0.009)
Fraction of new students who achieved Level 4 in Grade 3	0.049*** (0.008)	0.243*** (0.043)	-0.029** (0.014)	0.067*** (0.009)	0.062*** (0.015)
Panel B (IV)					
Fraction of current classmates who achieved Level 1 in Grade 6	-0.534 (0.570)	0.393 (0.887)	-1.434** (0.654)	-0.335 (0.670)	1.074 (1.063)
Fraction of current classmates who achieved Level 3 in Grade 6	0.217 (0.284)	-1.102 (0.693)	0.16 (0.372)	0.332 (0.308)	0.497 (0.368)
Fraction of current classmates who achieved Level 4 in Grade 6	0.568*** (0.184)	3.515** (1.433)	-0.185 (0.454)	0.671*** (0.217)	0.843*** (0.248)
Number of observations			199,717		

Note: Standard errors clustered at school level. *** p-value<0.01, ** p-value <0.05, * p-value<0.10. The sample comprises all incumbent students in Grade 3 for three years: 2008-2010. All regressions include school, year and school-by-year fixed effects. The coefficients in the first colum represent the effect of entry of new students with corresponding level of achievement on all incumbent students independent of their achievement level in Grade 3. Reduced form estimates are presented in Panel A when fraction of new students at different levels of achievement enters directly into the estimating equation. Panel B presents the coefficients of the 2SLS procedure when fraction of all classmates in Grade 6 at a specific level of achievement is instrumented with the fraction of new students only at that same level of achievement.

Table 8. Effect of the classmates on gain in test score for incumbent students with different achievement levels

	Effect on Grade 6 test score of incumbnet student if own test score in Grade 3 is:			
	Level 1	Level 2	Level 3	Level 4
Panel A				
Fraction of current classmates who achieved levels 1&2 in Grade 6	-1.410***	-1.482***	-1.106***	-0.623***
	(0.026)	(0.015)	(0.012)	(0.022)
Fraction of current classmates who achieved level 4 in Grade 6	0.204***	0.417***	1.066***	1.171***
	(0.064)	(0.026)	(0.017)	(0.024)
Panel B (Reduced form)				
Fraction of new students who achieved Levels 1&2 in Grade 3	0.142***	-0.132***	-0.030**	-0.010
	(0.031)	(0.015)	(0.012)	(0.020)
Fraction of new students who achieved Level 4 in Grade 3	0.296***	0.009	0.117**	0.080***
	(0.074)	(0.027)	(0.019)	(0.028)
Panel C (IV)				
Fraction of current classmates who achieved Levels 1&2 in Grade 6	-0.645	-0.694*	-0.190	0.495
	(0.490)	(0.360)	(0.321)	(0.422)
Fraction of current classmates who achieved Level 4 in Grade 6	0.395***	0.471	1.043***	1.001***
	(0.118)	(0.594)	(0.328)	(0.325)
Number of observations	199,717			

Note: Standard errors clustered at school level. *** p-value<0.01, ** p-value <0.05, * p-value<0.10. The sample comprises all incumbent students in Grade 3 for three years: 2008-2010. All regressions include school, year and school-by-year fixed effects. Panel A demonstrates coefficients from the fixed effects regression. Reduced form estimates are presented in Panel B when fraction of new students at different levels of achievement enters directly into the estimating equation. Panel C presents the coefficients of the 2SLS procedure when fraction of all classmates in Grade 6 at a specific level of achievement is instrumented with the fraction of new students only at that same level of achievement.

Table 9. Estimated effect of alternative classroom assignment on math test score gains

Panel A. Change from 20 percent level 1&2, 60 percent in level 3 and 20 percent level 4 **(20:60:20)** to 60 percent of level 1&2, 30 percent of level 3 and 10 percent of level 4 **(60:30:10)**

Level 1	0.0272
Level 2	-0.0528
Level 3	-0.0237
Level 4	-0.008

Panel B. Change from 20 percent level 1&2, 60 percent level 3 and 20 percent level 4 **(20:60:20)** to 10 percent of levels 1&2, 30 percent of level 3 and 60 percent of level 4 **(10:30:60)**

Level 1	0.1042
Level 2	0.0132
Level 3	0.0498
Level 4	0.032

Panel C. Change from 20 percent level 1&2, 60 percent level 3 and 20 percent level 4 (**20:60:20**) to 5 percent in levels 1 and 2 , 90 percent in level 3 and 5 percent in level 4 **(5:90:5)**

Level 1	-0.0657
Level 2	0.0198
Level 3	-0.01305
Level 4	-0.012

Note: Estimated effects represent the coefficients from Table 5 multiplied by the change in shares of peers at levels 1&2 and level 4.

Table 10. Effect of class heterogeneity on math test score gain of incumbent students

	Fixed effects	IV	Reduced form	IV			
				Level of achievement of incumbent student in G3			
				Level 1	Level 2	Level 3	Level 4
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Panel A							
Standard deviation of G6 math test score	0.115***	0.925		0.224	0.491	1.120*	1.652**
	(0.029)	(0.664)		(0.743)	(0.667)	(0.662)	(0.739)
Standard deviation of G3 math test score of new students			0.018				
			(0.011)				
Panel B							
Standard deviation of G6 math test score	0.141***	0.752		0.735	0.144	0.971	1.460**
	(0.03)	(0.630)		(0.848)	(0.653)	(0.645)	(0.692)
Average G6 math test score in class	0.061***	0.477**		0.367	0.658***	0.655***	0.707***
	(0.014)	(0.209)		(0.246)	(0.161)	(0.160)	(0.174)
Standard deviation of G3 math test score of new students in class			0.025**				
			(0.011)				
School, year and school-by-year fixed effects	Yes	Yes	Yes	Yes			
Number of observations	199,717	199,717	199,717	199,717			

Note: Standard errors clustered at school level. *** p-value<0.01, ** p-value <0.05, * p-value<0.10. The sample comprises all incumbent students in Grade 6 for three years: 2008-2010. The dependent variable is EQAO math test score in Grade 6. Standard deviation of Grade 6 math test score is defined as standard deviation over the test scores for all students in class except student i . Panel A presents results with standard deviation as the only peer variable in equation; Panel B includes both the average test score and standard deviation of test scores of classmates. Column (1) presents estimates of fixed effects specification; column (2) instrumental variables estimates when standard deviation of Grade 6 test scores is instrumented with the standard deviation of Grade 3 test scores of only new students in class. Column (3) is a reduced form spesification where standard deviation of Grade 3 test score of new students enters directly into equation. Columns (4)-(7) demostrate the effect of class heterogeneity measure by the standard deviation of test scores on incumbent students with different intial level of achievement (Grade 3 math test score).