Peer Effects in the Classroom: Evidence from New Peers

Margarita Pivovarova*

Preliminary and Incomplete

August 1, 2012

Abstract

Peer effects in education, their magnitude and structure is an open question among economists and policy-makers. This paper attempts to address the issue using administrative data on four cohorts of Ontario elementary school students to estimate the magnitude and structure of peer effects in the classroom. I use data on test scores from province-wide assessment of mathematics, reading and writing abilities of all students in public schools which takes place at the end of Grade 3 and then again at the end of Grade 6. I use direct measure of peer ability - test score in Grade 3 - which is immune to reflection problem because it is observed before students changed school. I use variation in the average ability of new peers (those who entered school between Grade 3 and Grade 6) as exogenous source of variation to identify peer effects at a classroom level. To identify the direction and magnitude of educational spillovers, I estimate a fixed effects model which accounts for non-random sorting of students into schools over time. However, this still leaves non-random allocation of students into classroom within school a problem. I claim that using a sample of schools with only one Grade 6 class and simultaneously accounting for school-cohort fixed effects can be considered a quasi-random allocation of new peers. I find that peers' ability matters for own student's achievement and the magnitude of these spillovera is not trivial. I also find that the structure of peer effects is non-linear. I test various model of heterogeneous peer effect and show which models of peer effects are supported by the data.

1 Introduction

Peer effects in education are of interest to parents, policy-makers and researchers alike. While recent literature has provided credible estimates of peer effects in elementary and

*PhD Candidate, Department of Economics, University of Toronto. e-mail: rita.pivovarova@utoronto.ca.
I am indebted to my supervisor, Dwayne Benjamin, for his patience and support. I also would like to thank Gustavo Bobonis for his helpful suggestions and discussion. I thank EQAO for providing me with the administrative data. Michael Kozlow of EQAO was the most helpful in accessing the data.

high education, these estimates are mixed and the structure of peer effects remains an open question.

In this paper, I use test scores data of elementary school students in the largest Canadian province of Ontario to estimate peer effects in the classroom. I also provide evidence against the linear-in-means model of peer effects and show which models of peer interaction are more consistent with the data.

Elementary school environment seem to be an appropriate context to study peer effects. Indeed, it is hard to find another non-experimental setting when individuals are forced to spend substantial number of hours every day for a prolonged period of time. Moreover, these individuals are subject to common shocks because in elementary school students spend time in the same classroom and receive instruction from the same teacher. Such non-experimental setting has a number of advantages over the controlled environment when individuals are randomly assigned to their peer groups. Classes in elementary schools are relatively large, and there is a choice of peers to form smaller groups within classes without being forced to do so. At the same time everyone in a class is a peer because of the close proximity to each other on everyday basis.

Much of the literature on peer effects in elementary and secondary schools finds modest to large effects of peer background on own test scores (Hoxby (2000), Hoxby and Weingarth (2006), Lavy and Schlosser (2007), Vigdor (2006), Vigdor and Nechyba (2007), Hanushek, Kain, Markman and Rivkin (2003)). However, using a large panel dataset of elementary students, Burke and Sass (2008) find little evidence that the peer average background affects the average student's academic achievement; the authors, however, find a sizable non-linear spillovers. Also, Imberman, Kugler and Sacerdote (2011) find no effect on average, but some evidence of peer effects in the middle of the distribution. Within the higher education, many studies find positive but very modest in magnitude peer effects on GPA (Carrell, Fullerton and West (2009), Sacerdote (2001), Zimmerman (2003)). The majority of the studies focuses on the existence of peer effects with the exception of Hoxby and Weingarth (2006) and

Imberman, Kugler and Sacerdote (2012) for the US, Lavy, Paserman and Schlosser (2008) for Israel, and Duflo, Dupas and Kremer (2010) for Kenya which test different models of peer interactions.

This paper contributes to the discussion of peer effects in elementary education, but is different from the studies above in two ways. First, unlike the majority of studies, I observe test scores for students before they switched school which means that my measure of peer quality is a direct measure of ability and is an exogenous measure with respect to the current peers outcomes. In the absence of experimental data in studies of peer effects, the most common strategy has been to proxy for peer ability or behavior using preexisting measures such as race and gender (Hoxby and Weingarth (2006), Hoxby(2000), Lavy and Schlosser (2011)), subgroup reassignment (Angrist and Lang (2004), Hoxby and Weingarth (2006)), the presence of boys with feminine names (Figlio (2007)), the presence of children who had previously been retained (Lavy, Paserman, and Schlosser (2007)), or the presence of family problems (Carrell and Hoekstra (2010)). I overcome reflection problem using pre-determined measure of peer ability, and in that my approach is similar to Imberman, Kugler and Sacerdote (2012) who use pre-existing test scores for Katrina's evacuees in Louisiana schools and Lavy, Silva and Weinhardt (2012) for English secondary schools. Second, I am able to identify student's peers at the classroom level and exploit within-school between classroom variation in peers pre-determined abilities.

Using pre-determined measure of peer ability and within-school between classroom variation in average peer quality, I find sizable and significant peer effects. My findings imply that adding new peers who raise the average test score by 1 standard deviation raises student's own achievement by 0.11 standard deviation. This effect is different for students who are new to the school (switchers) and students who stayed in the same school through Grades 3 to 6. The effect for new students is substantially larger - for one standard deviation increase in the average peer test score own student's achievement raises by one third of a standard deviation. While linear-in-means effects are statistically significant, the linear-in-means model is not supported by the data. The estimates of the heterogenous model of peer effects imply that the structure of peer interactions is non-linear with all students benefiting from the higher share of high-achieving peers but the magnitude of the effect depends on the student own ability. I find evidence in favour of the single-crossing model - an increase in the average ability of peers benefits high-achieving students more than low-achieving students. I do not find support for the other models of peer interactions - invidious comparison and boutique or tracking models. Rejection of boutique model implies that grouping students by ability into classes does not increase the aggregate level of achievement. While tracking by ability is a common practice in secondary schools, it does not seem to be an optimal mechanism for raising academic achievement in elementary school.

The rest of the paper is organized as follows. Next section provides brief review of the relevant literature. In section 3 I describe the data and provide background information about elementary education in Ontario. I explain challenges associated with estimation of peer effects and propose identification strategy in section 4. Empirical results are presented in section 5 separately for linear-in-means and heterogeneous model of peer interactions. Section 6 concludes with policy implications and suggestions for future work.

2 Peer Effects in the Literature

Recently the literature transitioned from just obtaining credible estimates of peer effects towards the tests of the structure of those effects. The structure of peer effects seems to be important as it allows to distinguish between various models of peer effects. The second related strand of the literature explores the various channels of peer effects in school.[*to be updated*]

3 Data

3.1 Education in Ontario

The public school system in Ontario is similar to other Canadian jurisdictions and the United States. The system consists of 72 school boards including English Public, English Catholic, French Public and French Catholic boards. Many of the school boards correspond to city boundaries in the populated areas (city of Toronto, for example). French schools represents about 10% of the elementary enrollment.[1] School boards are required to admit all students who or whose parents or guardians reside in the school section.[2] Elementary schools include Grades from 1 to 8 plus Kindergarten while secondary schools comprise Grades 9 to 12. The Education Quality and Accountability Office (EQAO) administers a number of province-wide tests in all publicly funded schools. The tests are based on the Ontario curriculum and are conducted every year at key stages of students' development. The assessments include tests of mathematics, readings and writing abilities in Grades 3 and 6, mathematics in Grade 9, and literacy test in Grade 10. The testing program in Ontario was created in 1996 and since then has gained support among educators [3]. The program is said to enable school boards to develop improvement plans, and although the EQAO opposes the use of the tests results for official school ranking, they are used by Fraser Institute in school annual report card ratings [4].

3.2 Data sources

The data for this study were obtained from the EQAO and consist of three data sets that were linked together as explained in Appendix. The first data set consists of records for all students in grade 6 for 2008, 2009 and 2010 school years. For these students, I know the

[1]Descriptive statistics by school board is provided in Appendix.

[2]Education Act,R.S.O., 1990, Chapter E2. There are additional requirements to attend Roman Catholic School and French-language School Board as outlined in the Education Act, 1990.

[3]See The Power of Ontario's Provincial Testing Program brochure which provides the results of the survey of principal, teachers and parents

[4]Report Card on Ontario's Elementary Schools 2011, Fraser Institute Studies in Education Policy

results of mathematics, reading and writing tests, whether student was excused from writing the test, whether student is in English as Second Language program (ESL), whether student has learning difficulties, gender, date when entered current school, whether student has been born in Canada and whether s/he learned English/French at home. For all students who sit the Grade 3 test, I know the results of their prior mathematics, reading and writing test scores. This data set allows me to construct class and school level variables used in quantitative analysis.

The second data set is an aggregated data for each school in Ontario with percentage of students who sit each of the three tests in Grade 6, percentage of male and female students, percentage of students who scored above and below provincial standards, percentage of students born in Canada. For each school I know the name of the school and postal address and can identify it as a rural or urban school and the school board to which school belongs.

The third data set is a file which contains characteristics of school neighborhood identified by the three first digits of the postal code, also known as Forward Sortation Area (FSA). These characteristics include median and mean household income, percentage of residents with university degree, percentage of recent immigrants, share of households living at poverty line, percentage of single parent families. Neighborhood characteristics were constructed by the EQAO from Canadian Census 2006. The characteristics of the community while linked to each school in fact represent the average characteristics of the students' residential neighborhood and thus serve as reasonable proxy for student's socio-economic status. [5]

I combined all three data sets into one aggregate file that contains information about each student, his/her classroom, school and school and neighborhood characteristics. The resulting file is a two-level panel: three year panel of schools and two year panel of students' test

[5]It is true that postal code might not reflect the student's neighborhood but rather the neighborhood where student actually goes to school. However, sorting happens later when students move from middle to high school. Secondary public schools in Ontario allow for optional attendance when a student may attend a school outside his/her catchment area provided there is enough space in that school. This is not the case for elementary schools.

score data.

As has been already mentioned, students in Ontario write province-wide tests of their reading, writing and math skills at key stages in their elementary and secondary education. At the end of primary (Grade 3) and junior (Grade 6) division both literacy and math are tested, at the end of Grade 9 students take a mathematics test, and at the end of Grade 10, students' literacy skills are tested. In this paper, I only use results of the tests in elementary school, at the end of Grade 3 and at the end of Grade 6. There is a number of reasons for doing that. The most important one is that the structure of the data does not allow me to track students from elementary to secondary school and as a result I cannot match tests scores in Grade 9 and 10 to test scores in Grade 3 and 6. The second reason is that the results of assessments in Grade 3 and Grade 6 are more comparable and consistent across years since they test students' knowledge in the same subjects - mathematics, reading and writing. The third reason is a methodological one. I use variation in the ability of incoming students to identify peer effects and argue that conditional on school-year fixed effect, assignment of students into classes is random. This assumption is more likely to hold for elementary school when students are placed in classes based on class size capacity, school principal's and teachers' considerations and on observed student's characteristics[6] Also, elementary school students spend most of their school time with their classmates for all subjects and with the same home teacher. [7] This way, the peer group for an elementary school student is clearly defined as his/her classmates. This is not the case for secondary school. In secondary school, students choose subjects and indirectly choose their peer group; to some extent, students sort themselves into subjects based on both observed and unobserved characteristics. The classmates no longer represent the peer group as students do not take the same courses with

[6] In informal talks with school principals, they almost uniformly confirmed that the placement of new students into classes is random and is based primarily on class size. Also, special needs of the students are taken into account, such as learning difficulties, English as a second language. Immigrant students are often placed in classes with students who speak the same language provided that the class size condition is met. Gender balance is also an important consideration when students are grouped into classes.

[7] As a rule, in Ontario elementary schools, all three subjects – mathematics, reading and writing are taught by the same teacher.

those who formally belong to the same classroom.

The results of the test scores in Grades 3 and 6 for all three subjects are reported on a scale from 1 to 4. These levels correspond to the standard system of evaluation of students in Ontario schools according to the Ontario Curriculum. In general, the results of the EQAO test provide a snapshot of a student's achievement relative to the provincial standards. Level 3 is the threshold level when student's achievement meets the provincial standards. A brief description of the levels and corresponding percentage score is given in the chart below:

Level	Provincial Standard	% Marks
Level 4	The student has demonstrated the required knowledge and skills. Achievement exceeds the provincial standards.	80-100
Level 3	The student has demonstrated most of the required knowledge and skills. Achievement meets the provincial standards.	70-79
Level 2	The student has demonstrated some of the required knowledge and skills. Achievement approaches the provincial standard.	60-69
Level 1	The student has demonstrated some of the required knowledge and skills in limited ways. Achievement falls much below the provincial standard.	50-59
0	The student has not demonstrated the required knowledge and skills. Extensive remediation is required.	Below 50

The discrete nature of test score presents a challenge for interpreting and comparing results. In quantitative analysis I predominantly use the level system of test scores especially when I classify students as high or low achievers. I also converted the four-level test score into the standardized score with mean zero and standard deviation equal to 1 to facilitate comparison of my results to other studies.

3.3 Sample restrictions

I have to impose a number of standard restrictions on my sample as well as additional restrictions specific to my data. First, I remove all observations with missing information about test scores. For mathematics, this reduces the sample size by 14%, from 429947 to 368769 observations for individual students. Additionally, because I was not able to match all schools to their neighborhoods by three digits postal code, I have to drop 6415 observations, or about 1% of original sample. However, given that most of the unmatched schools are school with less than 10 students in Grade 6, I would have dropped them from my analysis for another reason. Conventionally, in the peer effects in education literature, classes and school with small number of students are dropped from analysis. For observations with missing data on students characteristics, I create dummy variables indicating missing values and keep observations in my sample. I have a total of 26116 classrooms in my sample, 1160 of them only have one student in Grade 6. I drop such classrooms from analysis which reduces my sample further by 0.05% to 367109 individual observations in 24456 classrooms in unbalanced panel of 3467 schools for three years, for a total of 10417 school-year data points.

Another restriction imposed by the nature of the data is that even though I have data for five school years, I only use three cohorts of students - those who were in Grade 6 in 2007/2008, 2008/2009 and 2009/2010 school years as only for these three cohorts the data allow matching of records in Grade 3 and Grade 6.

For several regressions, I use sample with peers for whom I do not know test score in grade 3, and I use their test scores in grade 6 instead. This mostly concerns estimation with immigrant children as they constitutes the sizeable fraction of new peers who arrive to school between grade 3 and grade 6 outside Canada and hence do not have records of their past test scores. Another group of new peers for whom test scores are not available are those who moved from another province. One limitation of the data is that I cannot identify why the test score in Grade 3 is not available [8]. Instead, I provide summary statistics comparing students with missing test scores to their peers with test scores available for both grades. I discuss the findings in the next section.

[8] I run a series of OLS regression to estimate the propensity for not having Grade 3 test score and report results in Appendix Table 2

3.4 Old peers and new peers

To identify peer effects among Ontario elementary school students, I use two features of the data: short panel of schools and availability of test scores in Grade 3 and Grade 6 for large fraction of students in the sample. The first feature allows me to control for school specific time invariant unobserved heterogeneity, as well as for time variant effects and time trends. Using test score for two years, I can estimate a value added education production model of students achievement. Conventionally, the value added model includes student's own characteristics, parent, teacher and school inputs. When peer effects are estimated, production function also includes a measure of average peer quality. In this paper, the average peer quality is defined as average test score of peers in Grade 3 and the main challenge is to identify the relevant peer group. The structure of the data allows me to use the smallest level of aggregation – classroom – to identify peers.[9] Moreover, since I also know the date when a student enters a school, I can identify new peers for students who did not switch the school between Grade 3 and Grade 6. Specifically, I construct a group of new peers for every student in my data set as follows. For students who did not switch school in the observed period (I will call them "stayers" or "incumbents"), the following table defines new peers:

Cohort	Grade 3	Grade 4 and Grade 5	Grade 6
I	2004/2005	New peers are all students who entered in September 2006 and later	2007/2008
II	2005/2006	New peers are all students who entered in September 2007 and later	2008/2009
III	2006/2007	New peers are all students who entered in September 2008 and later	2009/2010

For students who switched schools between Grades 3 and 6 (I call them "switchers"), the new peer group is defined as all students in her/his class in Grade 6. In my sample, 36% of

[9] The accuracy of what researchers identify as a set of relevant peers is important for the ability to identify peer effect itself. The question has been raised in Carell *et al* (2008) among others.

all students for whom the test scores in both Grade 3 and Grade 6 are available switched school over the observed period.[10] Table 1 provides descriptive statistics for a limited number of observed demographics variables for old and new peers. As evident from the table, switchers and stayers are different in many respects except for the shares of female students. On average, test scores in both Grade 3 and Grade 6 for switchers are lower and this difference is statistically significant. Switchers are more likely to learn English as a second language (ESL), be born outside Canada and not learn English at home. It is also interesting to look at the observed characteristics of students for whom test scores in Grade 3 are not available. The descriptive statistics for those students as well for those for whom I do not have test score in Grade 6 is provided in columns (3)-(6) of Table 1. I deal with new peers for whom I do not have test score in Grade 3, in section [].

4 Identification Strategy

The identification of peer effects is a well known challenge as it requires dealing with at least three econometric problems summarized by Manski (1993)[11] and being referred to as contextual and correlated effects and reflection problem.

Reflection, or simultaneity problem stems from the reciprocal nature of peer interactions as it is reasonable to assume that the outcome of a peer group is affected by the outcome of each member of that group. In this case, the use of contemporaneous or lagged outcomes of peer group is problematic. To overcome the reflection problem, researchers use background characteristics of peer group such as gender, ethnicity, parental schooling to proxy for ability measure and to identify peer effect on students' outcomes. One limitation of these measures of peer quality is that they do not directly quantify peers ability. Hoxby and Weingarth (2006) show that when peers ability is properly accounted for, then race, ethnicity, income

[10]For comparison, Lavy, Silva and Weinhardt (2012) report that more than 80% of students in English secondary school face new peers.

[11]See also Moffitt (2001) and Brock and Durlauf (2001) for discussion about the challenges associated with identification of peer effects

and parental education have no effect on students' outcome beyond peers ability. The advantage of the EQAO data is that I can construct a direct measure of peer ability which is immune to reflection problem. I use Grade 3 test score of new peers which is not affected by student's peers in Grade 6. This is a direct measure of peers ability and it reflects a student's academic achievement relative to the uniform provincial standards. This measure is also exogenous to student's own current achievement and is not subject to simultaneity problem. The two other studies that use a similar measure of peers ability is Lavy, Silva and Weinhardt (2012) for English secondary schools and Imberman, Kugler and Sacerdote (2012) for students in Louisiana.

The second challenge is selection problem, or sorting of students across schools and neighborhoods. When random assignment of students to peer groups is not possible, the most popular method is to exploit within-school or within-grade variation in fixed characteristics of peers across cohorts. Other methods include within student variation, instrumental variables and students' reassignment. These methods provide a solid ground for identification of peer effects at the school, grade or cohort level. However, there are empirical evidence that the higher the aggregation level of peer group, the smaller are peer effects which suggests that these effects operate within a smaller peer group. For instance, Burke and Sass (2011) find significant peer effects only at the classroom level, and not at the grade level. In this study I am able to match students by classrooms within school and identify all new and old peers for every incumbent student in class. As a result, I can improve on existing studies by exploiting variation in peer characteristics between classes controlling for school and cohort unobserved and observed heterogeneity. The identification of peer effects in this setting comes from within-school between classroom variation in peer quality.

The baseline model in this paper is a standard value added model of education production function with linear-in-means peer effects augmented with school, year and school-by-year fixed effects:

$$Y^6_{icst} = \alpha Y^3_{icst} + \beta_1 Y^3_{\text{New Peers},icst} + \theta X_{icst} + \bar{X}_{-icst} + \gamma S + \delta N + \lambda t + \rho S \times t + \varepsilon_{icst}$$

where Y^6_{icnst} is an outcome of interest - Grade 6 test score - for student i in classroom c in school s at year t. $Y^3_{\text{New Peers},icst}$ captures peer effects and represents average test score of peer group in Grade 3, X is a set of individual controls, $\bar{X}_{-icst}$ are average background characteristics of peer group, S is a vector of time-invariant school characteristics, and N is a set of static neighborhood controls. Time fixed effects, t can be thought of as cohort effects. School fixed effects capture time-invariant between school idiosyncratic differences, and school-by-year fixed effects soak up unobserved time-variant heterogeneity among schools. Regression also includes student i's own lagged test score in the same subject in Grade 3, $Y^3_{ics,t-3}$. Finally, ε_{icst} is an error term. The error term can be further decomposed as $\varepsilon_{icst} =$. Individual disturbances are purged by taking the first difference of test scores for each individual student, $Y^6 - Y^3$. School fixed effect and cohort fixed effects account for all static unobservables at school level and cohort specific differences which might be correlated with the peer variable. The dynamic component of the error term is accounted for by inclusion of the school-year fixed effects. The remaining part of the error term - any changes in peer composition at classroom level, or teacher specific effects which are correlated with peer variable - presents a threat to identification. I explain how I deal with this problem in section [5.5].

In this specification of education production function, coefficient of $Y^3_{\text{New Peers},icst}$ represents the net effect of student i's peers on own achievement. This effect includes both the direct effect from peer to peer interactions plus indirect effect through the changes in teacher's instruction caused by the changes in the classroom composition. Also, the effect of the average peers' score is a net effect of new peers and "lost" peers. In the data I can not identify students who left the school after Grade 3, meaning that I cannot identify "lost" peers. However, in regression for stayers, I explicitly control for quality of "old" peers using their average Grade 3 test score.

The fixed effect model solve problem with selection between schools but is not helpful if students are assigned to classes in a systematic way. While the data I use do not come from an experiment, nor can it be considered a natural experiment, I argue that the assignment of new students to classrooms within school can be considered as good as random conditional on observed and unobserved school characteristics. Table 2 provides correlations between qualitative characteristics of old and new peers in a classroom conditional on observed school and neighborhood variables. The first immediate observation lends support to the identification strategy which treats year-to-year or grade-to-grade variation in gender composition as exogenous. Out of seven correlation for student's characteristics available in my data, shares of girls among old and new peers seem to be unrelated. Given that the majority of the public schools consider the gender balance in the classroom to be a priority, or at least try to avoid skewed gender ratios, this result seem to be plausible. The second extreme value in the table is the correlation between shares of English as a second language (ESL) learners. The fact that ESL students are placed in the same classroom with other students who learn English as a second language is not surprising. In informal talks with school principals they confirmed that they try to group English as a second language learners into same classrooms in order to efficiently use limited resources of ESL teachers. For the identification strategy in this paper to be valid, placement of new students into classes between Grade 3 and Grade 6 should be random conditional on the observed characteristics for the pool of all new students in school. If students are placed into classrooms in a systematic way – for instance, they are grouped by abilities based on their prior achievement, then I should observe strong correlation between shares of students in the same level of achievement for old and new peers. As shown in Table 2, this correlation is rather small – 0.17 for the low ability students and 0.23 for high ability students. The opposite story might be that school principals try to balance students with low and high abilities, and in this case I should observe strong negative correlation between shares of old and new peers in the opposite levels of achievement. Again, I do not find support for such balancing in the data - while the associated correlations are negative, these are small – from -0.067 to 0.08. The average ability of stayers and switchers are not strongly related either - the correlation between mean level of math test score of the incoming students and incumbent students is only 0.26 conditional on school and neighborhood variables.[12] Another evidence in favour of plausibly random assignment of students in classes is no difference in standard deviation of test scores at a class and school level. As noted in Vigdor and Nechyba (2004), if there is ability tracking in schools, then the standard deviation of test scores at class level should be much smaller than standard deviation at school level. In Ontario data, the standard deviation of math test scores at class level is 0.67 and at school level 0.72. These numbers suggest that there is little evidence of ability tracking at class level within elementary schools. This, however, does not imply that students in all schools are assigned to classes randomly. [13]

5 Results

5.1 Linear-in-means model

The main results in this section are based on the estimation using mathematics test scores. In Appendix Tables A1 and A2 I provide results for reading and writing test scores. While the magnitude of the coefficients is slightly different, all the findings discussed below are qualitatively the same. Table 3 shows the results of the estimation of equation (1) for various samples. In column (1), I present OLS estimates with individual controls only – gender, ESL status, Canadian born and first language. The coefficient on peer variable is positive and statistically significant implying that on average higher quality of new peers in class is associated with slight improvement in test score from grade 3 to grade 6. Since OLS estimates are most likely subject to selection bias and do not represent the causal

[12]In Appendix, I provide the results of the survey conducted among elementary school principals in Ontario. These results support the assumption of students being placed into classes not based on their ability but rather on class size capacity, making ability grouping or mixing unsystematic. These results concern students who are new to the school. The assignment of students into classes happens at the end of the school year, usually in May. In the informal talks with school principals they explained that such assignment is made based on the number of individual student's characteristics with the final goal of balanced classroom across all of these characteristics - gender, ability, social skills, newcomer status, student's special needs.

[13]Vigdor and Nechyba (2004) consider the difference of 0.987 at school level and 0.432 at class level to be suggesting of at least some degree of ability stratification.

effect of the average peer quality on student's own achievement, I move on to estimating equation (1) with fixed effects.In Column (2), I include year, school and year-school fixed effects as well as the neighborhood controls. Columns (1) and (2) presents estimates for the entire sample. The effect of peer quality measured by the average test score of incoming students (switchers), is significant, positive and large in magnitude.[14] Interestingly enough, without school-year fixed effect, the coefficient of mean peer achievement drops two-fold when I control for the neighborhood observed characteristics, but when school-year fixed effect are added, the coefficient swings back to its original magnitude[15]. One way to explain the swing is to think about what do the neighborhood and school fixed effects capture in the regression. School-year fixed effects soak all the unobserved time-variant differences between schools and thus account for sorting of students and parents across schools over time, year fixed effects absorb time trends and school fixed effects account for unobserved time-invariant heterogeneity across schools. Neighborhood characteristics in my data do not vary over time and capture the neighborhood characteristics of students as if they lived there in 2006. Dropping them from the analysis while controlling for school fixed effects and time trends does not change the results of the estimation. To some extent, the above observation implies that (1) placement of new students into classes conditional on the observed school and neighborhood characteristics is random since school-year fixed effects do not change the estimate, and (2) parents sort themselves into neighborhoods creating non-random variation in average peer quality between schools.

The next step is to explore whether aggregate effect suppresses any differences between stayers and switchers. According to results in columns (3) and (4) this is indeed the case - effect for switchers is larger in magnitude and statistically significant. One explanation comes to mind - switchers might be more sensitive to changes in peer composition because

[14]I estimate the same regression using standardized test scores and report results in Appendix. While standardized results are easier to interpret, I report them in Appendix because in models with heterogeneous effects as well as in some of the linear-in-means models, I prefer to use levels of achievement rather than standardized scores. The results with standardized test scores are quantitatively the same, with patterns across different samples repeating patterns for unstandardized scores.

[15]Results are presented in Appendix

they experience not only the new peers in class but they are also exposed to the new school environment, and the observed effect is a compound impact of peers, school and neighborhood effects.[16] The peer effect for stayers is also significant, but smaller in magnitude. The estimated effect can be interpreted as follows: adding new peers who raise the average achievement of peers by 0.5 raises own student's achievement by 0.07, or by 10% of standard deviation of the test score. This number if three times larger for switchers – for the same change in the average peer achievement, own test score for switchers raises by 0.2. For stayers, it is two times smaller - 0.03. These results are not immediately interpretable because my measure of test scores is discrete, and not continuous as in other studies. So, translating into z-scores, the same results read as follows: adding new peers who raise average peer score by 1 standard deviation raises own student's achievement by 0.11 standard deviation, by 0.33 standard deviations for switchers and 0.07 standard deviation for stayers. However, as pointed out by Hoxby and Weingarth (2006), the average effect itself does not provide important policy implications as it only indicates that switching students with higher abilities across classrooms, raises or decreases the average test score in a classroom, but does not change the average level of achievement overall. I explore the differential impact of peer average quality on students at different levels of achievement in the next section.

Moving on to potential gender heterogeneity in responses to average peer quality, I estimate separate regressions for boys and girls and report results in Columns (5) and (6) of Table 3. While the difference between peer effects in for boys and girls is statistically significant, relative to the magnitude of the coefficient it is not important. Overall, it does not seem that the average effect of the new peers differs by gender. There is no gender difference in response among stayers and switchers as well.[17]

[16] There are no examples in the literature which compare the effects on switchers and stayers. Imberman, Kugler and Sacerdote (2012) report results for stayers, or incumbent students only, so my results are not directly comparable. Also, Lavy, Silva and Weinhardt (2012) use similar approach by identifying new and old peers but they do not report results separately because about 80% of students in English schools experience new peers when transition to secondary school. In Appendix, I report result for all regressions in this paper separately for switchers and discuss them briefly in section [].

[17] Results are available upon request.

There is no reason to expect that students at different levels of achievement should respond to changes in peer composition in the same way. Columns (7)-(10) show that this is indeed the case for the sample of Ontario elementary school students. The positive effect of the average peer quality is decreasing in own ability as measured by achievement in grade 3. Having more highly achieving peers is more beneficial for students whose own achievement is below provincial standard. The same pattern holds for stayers and switchers with effect on switchers being on average higher than for stayers. Overall, the results of the linear-in-means model shows that there is heterogeneity in how students in elementary schools respond to peer quality.

Even more interesting is to see whether the linear-in-means model holds among the elementary students in Ontario. The recent advancement in the peer effect literature exposed the For instance, Lavy, Silva and Weinhardt (2012) find heterogeneous peer effects ranging from negative effect from bad peers for students at the bottom of the ability distribution to no effect from good peers, and Lavy, Passerman and Schlosser (2007) find that high achieving students benefit from the presence of high achieving peers, while there is no effect for average students. Bett and Zau (2004) provide evidence that an average student is hurt more by low achieving peers than s/he is helped by high achieving peers. Hoxby and Weingarth (2006), Imberman, Kugler and Sacerdote (2012) among others reject linear-in-means models in favour of alternative models of the structure of peer effects. In section 5.2, I discuss which models of peer effects are supported by the data in this paper. In the remainder of this section, I will join the pool of researchers above and present evidence against the linear-in-means model. Table 4 shows the results of linear-in-means model for peer achievement when own level of achievement is interacted with the average peer achievement. Since my measure of test scores does not allow me to use the finer grid for test scores distribution (for instance, deciles as in Hoxby and Weingarth 2006, or bottom fifth and top fifth percentiles as in Lavy, Silva and Weinhardt (2012)), I use the EQAO four-level classification to distinguish between low achieving and high achieving peers. I estimate the following specification of equation (1):

$$Y^6_{icst} = \alpha Y^3_{icst} + \beta_1 Y^3_{\text{New Peers}} \times D^3_{L1,icst} + \beta_2 Y^3_{\text{New Peers}} \times D^3_{L2,icst} + \beta_3 Y^3_{\text{New Peers}} \times D^3_{L3,icst}$$
$$+\beta_4 Y^3_{\text{New Peers}} \times D^3_{L4,icst} + \theta X_i + \gamma S_i + \delta N_i + \lambda t + \rho S_i \times t + \varepsilon_{icst}$$

where subscripts $L1, L2, L3, L4$ denote corresponding levels of achievement with Level 1 being the lowest and Level 4 being the highest.This flexible specification allows the effect of the average peer achievement vary depending on student's own achievement. This econometric specification also provides a test for the validity of the linear-in-means model. Three panels of Table 4 present estimates for three different samples of elementary students - entire sample in panel A, stayers in panel B, and switchers in panel C. The coefficient estimates and F-statistics for joint equality of the coefficients imply that in all of these three cases the linear-in-means model is rejected. While the impact of peers ability on own test score is positive and statistically significant for students in all levels of achievement, this impact differs significantly. As before, it is more pronounced for switchers. Panel C shows, that among switchers, students who were at the highest level of achievement in Grade 3 benefit twice as much from the presence of highly achieving peers than those students who themselves were at the lowest level of achievement in Grade 3. Overall, the impact of highly achieving peers is positive for all students. At the same time, the average effect masks important difference between stayers and switchers, and between low- and high-achieving students. While the linear-in-means model also does not find support among the sample of stayers, the coefficient estimates do not exhibit the same pattern as for the entire sample and for switchers. Among incumbent students students at the lower end of ability distribution benefit from the increase in the average quality of new peers more than everyone else. Students who are at level 2 of provincial standards seem to benefit less than all other groups. For stayers, the average quality of old peers is included in all regressions. When the quality of the old peers is omitted, results do not change but the magnitude of the coefficients goes up by about 10% of the original magnitude.

Lastly, how do the estimates of peer effects for linear-in-means model in this paper compare to these in other studies? Given the lack of prior achievement data for peers, the majority of studies estimate peer effects in education using variation in peer background characteristics, and not the direct measure of ability or prior achievement: gender (Hoxby (2000),Whitmore (2005), Lavy and Schlosser (2007)), race (Hoxby (2000), Hoxby and Weingarth (2006)), participants of government programs ("Metco" busing program in Angrist and Lang (2004)). Sacerdote (2011) provides comparison of standardized results expressed as effect of 1 point increase in the average peer score on own test score across recent studies of peer effects in primary and secondary education. The estimates for math test score range from 6.8 (in Hoxby (2000) using gender variation) to -0.12 (in Vigdor and Nechyba (2007) using school, year and teacher fixed effects). My estimates of the average peer effect are modest-to-small compared to other studies - I find that a one standard deviation increase in peer mean math score raises own achievement by 0.11 standard deviation in the main specification, and by 0.07 standard deviation for stayers and 0.33 standard deviation for switchers. All of these effects are statistically significant.

5.2 Models of Heterogeneous Peer Effects

In this section, I explore the structure of peer effects, and estimate the model which incorporates all potential interactions between peer abilities and own achievement level. Unlike linear in means model which assumes that each student affects all of his peer identically, the flexible model of heterogenous effects allows this effect to differ depending on how different a student and his peers initially are. The econometric specification of heterogeneous effects contains 12 interactions as follows:

$$Y^6_{icst} = \alpha Y^3_{icst} + \sum_{j=1,3,4} \sum_{l=1}^{4} \beta_{jl} \text{Fraction}^j_{icst} \times D^l_{icst}$$
$$+ \theta X_i + \gamma S_i + \delta N_i + \lambda t + \varepsilon_{icst}$$

where Fraction^j_{icst} is a share of new peers of student i in class c in school s whose achievement level in Grade 3 is j. These shares are interacted with own level of achievement of student

i in Grade 3. Comparison between signs and magnitude of the coefficients β_{jl} allow me to test various models of peer effects which are described in Hoxby and Weingarth (2005) and Sacerdote (2011). All interactions with the share of new peers in Level 2 are omitted due to collinearity. The results of the estimation presented in Table 5 separately for stayers and switchers.

First column of Table 5 shows impact of peers of different ability levels on all students. The results are consistent across all three samples - increase in the share of high ability peers is beneficial for all students independent of the own level of achievement, and having relatively more peers whose achievement in Grade 3 was the lowest, negatively affects own achievement. As before, the magnitude of the effect for switchers is larger than for stayers.

Interpretation of all the coefficients is not as straightforward as in the previous regressions. All coefficient on interactions indicate the net effect on student's own test score when the fraction of peers in given level goes up *and* fraction of peers in Level 2 goes down. This is because the average level of peer group, class size and shares of peers at other levels are held constant. The results reveal considerable heterogeneity in the nature of effect by peer ability group and own achievement. Panel A of Table 5 shows results for the entire sample. One pattern is immediate - given the same level of average peer ability, and decreasing fraction of peers in Level 2 while increasing fraction of new peers who are at Level 1, has negative effect on all students except for those who themselves were at Level 1 in Grade 3. For those students effect is also negative but imprecisely estimated. Quite an opposite picture emerges when looking at the high achieving peers. For all students, independent on their own level of achievement in Grade 3, having more high achieving peers and less peers at level 2 is beneficial. The magnitude of the coefficient varies, but the equality of the effect for students in Level 1, 3 and 4 cannot be rejected, and only effect on students at level 2 is smaller.

For stayers, the effect of having more low achieving peers is also negative for all students, but not significant for those who were the lowest achieving (Level 2) and highest achieving

(Level 4) in Grade 3. For both switchers and stayers, these two group of students seem to be immune to changes in the proportion of low achieving peers. Again, switchers are more responsive to average quality of peers and their relative proportion in class - the magnitude of the coefficients is larger and they are more often significant compared to the sample of stayers.

The coefficient estimates are of interest not on their own, but in connection to various peer effects models that might be tested using the results of the estimation. The fully saturated model allows to test for weak and strong monotonicity, invidious comparison model, boutique or tracking and single crossing models.

I start with the tracking model as it provides clear policy implication for grouping students by abilities. Ability grouping or streaming is thought to be useful 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. [18] In this paper, the test for the tracking model of peer interactions would not be rejected if the increase in the share of peers from the same level of achievement as student's own has bigger 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 highest achieving students (Level 4) in all three samples, there is no evidence to support the tracking model for all other students. Previous studies by Hoxby and Weingarth (2006) and Imberman, Kugler and Sacerdote (2012) also found little support for the boutique model, while Burke and Sass (2011) provide evidence that tracking by abilities in elementary schools might be beneficial.

[18] Data from the survey of principals in Canadian secondary schools collected within PISA project in 2006 and 2009 imply that in the surveyed schools students are grouped by abilities for all subjects (10%) or for some subjects (72-75%). The exact question and the summary statistics of the responses by year are provided in Appendix.

The magnitude and statistical significance of the effect of having more high-score peers in class lend support to the Shining Light model. The Shining Light model implies that one bright student in class is enough to provide motivation and inspiration for others. Results for the entire sample and for switchers are consistent with the Shining Light model of peer interactions, while in the sample of stayers, low-achieving students do not seem to benefit from the boost in the share of high-scoring peers.

The next model of peer interactions that can be analyzed with the Ontario elementary schools data is the model of invidious comparison. According to the invidious comparison model, a student is harmed by the presence of higher achieving peers and is helped by the presence of low-achieving peers. In other words, if a Level 2 student experiences an increase in the share of peers whose Grade 3 test score was higher than the student's own score, then his/her own achievement drops down, while if the share of low achieving peers goes up, then his or her academic performance may improve. 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 are harmed by 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 or Level 3) is positive and significant for low-achieving students in the entire sample. The magnitude of the impact is large - increase in the share of Level 4 peers by 10% implies improvement in own test score in Grade 6 by 0.05 for Level 1 students. This is roughly the same effect as from an increase in the average achievement level of new peers by one standard deviation.

Finally, the model that has found support in previous studies of the structure of peer effects (Hoxby and Weingarth (2006), Imberman, Kugler and Sacerdote (2012)), is the single-crossing model, which is often referred to as monotonicity property of peer interactions. Under strong monotonicity, the positive impact from having high achieving peers is strongly increasing in own ability. To test this property, I make a pair-wise comparison of the coefficient estimates for adjacent levels of achievement. For instance, I compare whether estimates of the peer effect are statistically significantly higher for students whose own level of achievement is 4 than for those whose level is 3. Using the shares of new peers at four different levels of achievement, I compare coefficients β_{jl} and test whether $\beta_{jl} \geq \beta_{j,l-1}$. I only conduct this test for shares of high-achieving peers, whose academic score in Grade 3 was Level 3 and Level 4. Unlike a number of previous studies, I do not find evidence in favour of single-crossing model when I estimate the fully saturated model of peer interactions. For the entire sample of students, the effect of having more peers in Level 4 is the same for students whose own score is 3 or 4, but this effect statistically larger than for low achieving students - Level 2, and smaller than for the lowest scoring students - Level 1. For the sample of stayers, Level 3 students benefit from higher achieving peers even more than high achieving students themselves. Results for the low achievers are inconclusive since the effect of the share of high achieving peers for them is imprecisely estimated. Among the new students, or switchers, both group of high-scoring students (Levels 3 and 4) benefit from an increase in the share of high achieving peers. Moreover, they benefit more than low-scoring students, and this difference is statistically significant. The weaker version of monotonicity property – which is a flexible specification of linear-in-means model of peer effects – implies that the impact of the average peer ability is increasing in student's own ability not taking into account the shares of peers at different levels of ability distribution. The EQAO data are consistent with this weaker model of monotonicity. Referring back to Table 3, the weak monotonicity property holds for the sample of switchers - the peer effect is significantly greater for each consecutive level of achievement. It does not hold for the sample of stayers - while the effect of average peer ability is statistically the same for both Level 3 and Level 4 students and larger than the effect on Level 2 students, it is smaller than for the lowest scoring students. For the entire sample, with the exception of students whose own achievement is Level 2, the weak monotonicity property holds.

The overall findings imply that the structure of peer effects in elementary schools 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.

5.3 Sensitivity Analysis

In this section, I impose additional restrictions on my sample to make sure that the main result is not driven by the special cases of classrooms with only new peers or with only old peers. Table 3 in Appendix lists coefficient of fixed-effects estimation for five different samples and compares them to the main result in the first column. Overall, the results are in line with the main sample, the coefficients on peer variable are positive and statistically significant, but different in magnitude. There are two main observations. When only classes with new peers are taken into account, the peer effects rises three-fold. In such classes each student faces new peers from Grade 3 to Grade 6 because I do not observe those students in the same school in Grade 3. One way to think about the coefficient on peer variables is to think of it as a compound effect of peer ability and other background characteristics unobserved in my data set, such as parental education and race. For instance, racial composition of new peers might have a significant instantaneous impact but disappear in the long run, while ability might have a lasting effect over the years. Take for example results from North Carolina elementary schools in Vygdor and Nechyba (2004) who find a persistent effect of peer ability through Grades 5 to 8, but only instantaneous effect of classroom racial composition which completely disappears by Grade 8. If this is the case in Ontario elementary school, then the difference in the coefficient estimates for only new and only old peers seem to be plausible. The second interesting finding in Appendix Table 3 is that the smaller the class size the bigger the peer effect. This is again consistent with the literature when the effect is more likely to be found if the peer group is defined on a smaller scale. This can be thought as the diffusion of the peer effect in a relatively large group, such as school relative to a smaller group, a classroom. The observed effect in smaller classes is twice larger than for classes with more than 10 students [19].

5.4 Apparently Random Assignment

The random assignment of students into classroom is a doubtful feature of the data. More likely is that students are assigned to classes in such a way as to form heterogenous groups - by ability, gender and social skills. However, when school has only one class for a given grade, then school-year fixed effects also account for non-random sorting within school because two levels of sorting - school and class - would converge to one. In other words, the two components of the idiosyncratic error term in equation (1) are identical. Recall that the error term in equation (1) can be decomposed as follows: [Insert error term decomposition here]. In this section, I use a sample of school with only one Grade 6 class. In this case, school-year fixed effects absorb unobserved dynamic differences between schools, school fixed effects account for time-invariant differences; cohort effect is purged by the year fixed effects. The source of the apparently random variation in the average peer characteristics and ability is at a classrom level because in such setting each student randomly faces a different group of peers by construction of the peer group (as described in section []). Stayers face only new peers - those who entered school between start of Grade 4 and Grade 6. For the switchers peer group consists of all incoming students plus those who stayed in school since Grade 3. The sample consists of all schools with only one Grade 6 class. The identification of the average peer effects thus comes from the within class between students variation in average ability of peers.

Before I proceed to results of the estimation, it is worthwhile to compare students and school characteristics for two samples. The sample of schools with only one Grade 6 class comprises

[19] The median class size in Grade 6 in my sample is 18 students with the maximum number of students of 37. Overall, 30% of classes have less than 10 students and 3% have more than 30 students.

a large fraction of all schools in my sample - 852 out of 3395, or 25% of all schools, and accounts for 12% of all students with Grade 3 and Grade 6 test scores available. Descriptive statistics presented in Table 6. Two samples differ significantly along three characteristics: fraction of new peers, proportion of students born outside Canada and urban/rural location of the school. Only 62% of schools with one Grade 6 class belong to urban areas, while for schools with more than one Grade 6 class this proportion is 90%. Fraction of new peers and those who were born outside Canada is larger in the main sample. These three characteristics are correlated in a sense that one expects to observe higher peer turnover in urban areas compared to rural schools. Also, immigrants are usually settle in urban areas, which explain larger fraction of students born outside of Canada in a sample with predominantly urban schools. For the rest of the observed characteristics of students - test scores in Grades 3 and 6, fraction of female students and fraction of ESL students - there are no significant differences.[20]

Table 7 presents results of the estimation of equation (1) using the sample of students from school with one Grade 6 class only. Focusing on the coefficient of the average peer quality and moving from OLS estimation with individual controls to fixed effects model, the coefficient increases ten-fold and gains significance. The magnitude and sign of the average new peers quality is comparable to the result of the main sample. The puzzling observation is that moving from OLS in the main sample involves no change in the magnitude and significance of the coefficient, while for the sample with apparently random assignment, the OLS result indicates no peer effect at all.

5.5 Endogenous sorting into classrooms within school

In this section, I use instrumental variables technique to address the sorting of students into schools and evaluate the direction of the bias in the OLS estimates. According to results in Table 4 in Appendix, the selection story behind the data is as follows: the bias

[20] The difference in fractions and average test score is statistically significant at 1% level, however, the magnitude of the difference is small.

in the coefficient estimates is a negative one. The majority of the studies have found a positive bias in the OLS estimates which is reasonably explained by sorting of parents and students into schools and neighborhoods. When more able students are matched with the students of similar abilities and are enrolled into schools with better overall achievement records, then we are likely to find peer effects even if in fact there are no spillovers from one student to another. Exploiting instrumental variables and fixed effects strategies, a number of studies found that this is indeed the case and the OLS estimates of peer effects are upwardly biased. One important thing needs to be mentioned here. The majority of the studies rely on the grade-cohort variation which compares cohorts of students within the same school over years. School fixed effects and school specific time trends then take into account sorting across schools over time and the positive selection bias is a reasonable explanation. In my data, I exploit between-classes variation accounting for sorting across schools over time. The second selection issue that arises in this case is sorting between classrooms. If, for instance, students are grouped by abilities, then selection into classes is positive and together with positive sorting across school that would produce estimates of peer effects when in fact there are no spillovers. But if students are grouped into classes in such a way as to build heterogenous classrooms, then the changes in the average ability of classmates over time which are captured by the unobserved error term are systematically negatively correlated with student's own ability. If more able student is more likely to be matched with someone with lower abilities, then OLS regression would yield no peer effect while the spillovers are present. The resulting compound bias in the OLS estimates would be negative which is what I find when I use instrumental variables[21] .

to be updated

[21]This is a well know formula of statistical bias in OLS estimate due to omitted variables (this version is from Angrist and Pischke (2009, p.60): $\frac{Cov(Y_i,X_i)}{V(X_i)} = \rho + \gamma' \delta_{ZX}$ where γ is the OLS coefficient from regression of dependent variable Y on independent and omitted variables, and δ_{ZX} is the slope from regression of omitted variable Z on independent variable X.

6 Conclusion

This paper contributes to the discussion about the existence and magnitude of peer effects in elementary education and provides evidence in favor of sizable and non-linear spillovers among students in Ontario.

References

[1] Hoxby, C. (2000). Peer effects in the classsrom: Learning from gender and race variation. *NBER Working Paper 7867.*

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

[3] Imberman, S. A., Kugler, A. & Sacerdote, B. (2012). Katrina's children: Evidence on the structure of peer effects from hurricane evacuees. *American Economic Review*, forthcoming.

[4] 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.

[5] Carrel, S. E., Hoekstra, M.L. (2010). Externalities in the classroom: How children exposed to domestic violence affect everyone's kids.*American Economic Journal: Applied Economics*, 2(1), 211-228.

[6] Lavy, V., Schlosser, A. (2011). Mechanisms and impacts of gender peer effects at school. *American Economic Journal: Applied Economics*, April 2011.

Tables

Table 1. Summary Statistics

	Test scores in Grade 3 and 6 are observed		Test score in Grade 3 is not observed		Grade 3 and 6 test scores are not observed	
	Stayers	Switchers	Stayers	Switchers	Stayers	Switchers
Average math test score, Grade 3	2.82	2.75				
	(0.66)	(0.71)				
Average math test score, Grade 6	2.78	2.67	2.25	2.56		
	(0.76)	(0.80)	(0.90)	(0.90)		
Female	0.50	0.50	0.40	0.45	0.36	0.37
	(0.50)	(0.50)	(0.49)	(0.50)	(0.48)	(0.48)
English as a second language (ESL)	0.012	0.018	0.03	0.08	0.01	0.06
	(0.11)	(0.13)	(0.16)	(0.27)	(0.10)	(0.24)
Born outside Canada	0.06	0.14	0.13	0.44	0.06	0.35
	(0.23)	(0.34)	(0.34)	(0.50)	(0.24)	(0.48)
Did not learn English at home	0.14	0.26	0.21	0.43	0.13	0.40
	(0.35)	(0.44)	(0.40)	(0.49)	(0.34)	(0.49)
Number of observations	224,649	128,578	12,803	43,247	4,069	6,781

Note: Stayers are students who did not switch school between Grade 3 and Grade 6, switchers are students who moved to a new school after Grade 3 and before taking Grade 6 math test. Standard errors in parentheses.

Table 2: Correlation between New and Old Peers, by Classroom

	Dependent variable	Old Peers						
		Average Grade 3 test score	Fraction female students	Fraciton born outside Canada	Fraction ESL learners	Fraction did not learn English at home	Fraction of Level 1 students	Fraction of Level 4 students
New Peers	Average Grade 3 test score	0.257 (0.003)						
	Fraction female students		0.001 (0.003)					
	Fraciton born outside Canada			0.257 (0.004)				
	Fraction ESL learners				0.758 (0.002)			
	Fraction did not learn English at home					0.545 (0.002)		
	Fraction of Level 1 students						0.173 (0.011)	-0.067 (0.007)
	Fraction of Level 4 students						-0.081 (0.014)	0.231 (0.009)
	Number of observations	26116	26116	26116	26116	26116		

Note: Correlations for every classroom conditional on school and neighborhood characteristics: urban or rural school, Catholic school board, French language school board, average Grade 3 test score for old peers; median household income in the neighborhood, fraciton of lone parent familie, fraction of immigrant, fraciton of residents with university degree, fraction of low income families. Number of observation is the number of classrooms.

Table 3. Effect of the Average Quality of New Peers in Class on Test Scores

Dependent variable	Mathematics test score in Grade 6										
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
	All	All	All	Switchers	Stayers	Boys	Girls	Level 1	Level 2	Level 3	Level 4
Average Grade 3 mathematics test score of new peers in class	0.131***	0.077***	0.133***	0.383***	0.068	0.151***	0.113***	0.167***	0.163***	0.120***	0.105***
	(0.007)	(0.006)	(0.008)	(0.014)	(0.007)	(0.008)	(0.008)	(0.021)	(0.011)	(0.008)	(0.014)
Individual controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
FSA controls	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year fixed effects	No	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
School fixed effects	No	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
School-year fixed effects	No	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Number of observations	353227	353227	353227	128578	224649	177542	175703	10434	89824	212999	39970
R^2	0.298	0.309	0.310	0.346	0.291	0.307	0.314	0.043	0.048	0.046	0.046

Note: Standard errors clustered at school level. *** p-value<0.01, ** p-value <0.05, * p-value<0.10. In columns (1) - (3) sample includes all students for whom data on test score in Grade 3 and Grade 6 are available in the data. Column (4) only includes students who switched schools between grade 3 and Grade 6. Column (5) include students who did not switch schools over the observed period. Columns (8)-(11) present estimates for students at different levels of achievement. All regressions include individual controls (own test score in Grade 3, gender, Enlglish as Second Language learner, Canadian born, and whether student learned English at home), school controls - urban school, Catholic school board, school from Toronto Metropolitan Area; all regressions exept in column (1) include neighborhood controls - log of median household income, proportion of residents with university degree, some university, proportion of low income families, and proportion of recent immigrants. Regression in column (5) also include control for the average achievement of "old peers". Results with and without inclusion of this controls are quantitatively the same.

Table 4. Linear-in-means model of peers effects

	Average Grade 3 Peers Score x Level 1	Average Grade 3 Peers Score x Level 2	Average Grade 3 Peers Score* x Level 3	Average Grade 3 Peers Score* x Level 4
A. All				
Own math score in Grade 6	0.108***	0.086***	0.148***	0.173***
	(0.012)	(0.008)	(0.008)	(0.010)
Number of observations	353227			
F-test for joint equality	307.31***			
B. Stayers				
Own math score in Grade 6	0.095***	0.042***	0.077***	0.072***
	(0.014)	(0.009)	(0.007)	(0.010)
Number of observations	224649			
F-test of joint equality	172.20***			
C. Switchers				
Own math score in Grade 6	0.233***	0.290***	0.419***	0.505***
	(0.020)	(0.015)	(0.014)	(0.016)
Number of observations	128578			
F-test of joint equality	190.54***			

Note: Standard errors clustered at school level. *** p-value<0.01, ** p-value <0.05, * p-value<0.10. Coefficients are estimates of peer effects which are aloowed to vary depending on the student's own test score in Grade 3. Panel A uses the entire sample, Panel B presents estimates for students who did not switch school between Grade 3 and Grade 6 , and Panel C shows estimates for new students only.All regressions include set of individual controls (own test score in Grade 3, gender, Enlglish as Second Language learner, Canadian born, and whether student learned English at home), school controls - urban school, Catholic school board, school from Toronto Metropolitan Area; all regressions include neighborhood controls - log of median household income, proportion of residents with university degree, some university, proportion of low income families, and proportion of recent immigrants. Regression in Panel B also include control for the average achievement of "old peers".

Table 4. Heterogeneity of Peer Effects

	Level1 - Level 4	Own Grade 3 test score of incumbent student			
		Level 1	Level 2	Level 3	Level 4
			A. All		
Fraction of new peers in Level 1	-0.266***	-0.104	-0.419***	-0.221***	-0.155*
	(0.081)	(0.097)	(0.085)	(0.082)	(0.089)
Fraction of new peers in Level 3	0.180***	0.314***	0.116*	0.204***	0.177**
	(0.068)	(0.071)	(0.069)	(0.069)	(0.070)
Fraction of new peers in Level 4	0.376***	0.492*	0.245*	0.407***	0.412***
	(0.136)	(0.147)	(0.137)	(0.137)	(0.139)
Number of observations			353227		
			B. Stayers		
Fraction of new peers in Level 1	-0.226***	-0.076	-0.340***	-0.195**	-0.147
	(0.082)	(0.103)	(0.087)	(0.083)	(0.092)
Fraction of new peers in Level 3	0.157**	0.322***	0.112	0.178**	0.126*
	(0.070)	(0.074)	(0.071)	(0.071)	(0.071)
Fraction of new peers in Level 4	0.280**	0.474	0.208	0.316**	0.239*
	(0.140)	(0.153)	(0.141)	(0.140)	(0.142)
Number of observations			224649		
			C. Switchers		
Fraction of new peers in Level 1	-0.434***	-0.174	-0.641***	-0.402***	-0.243
	(0.139)	(0.170)	(0.142)	(0.144)	(0.186)
Fraction of new peers in Level 3	0.339***	0.205*	0.171	0.390***	0.522***
	(0.111)	(0.114)	(0.110)	(0.113)	(0.118)
Fraction of new peers in Level 4	0.759***	0.583**	0.472**	0.801***	0.821***
	(0.221)	(0.252)	(0.223)	(0.222)	(0.225)
Number of observations			128578		

Note: Standard errors clustered at school level. *** p-value<0.01, ** p-value <0.05, * p-value<0.10. Coefficients are estimates of peer effects which are allowed to vary depending on the student's own test score in Grade 3. Panel A uses the entire sample, Panel B presents estimates for students who did not switch school between Grade 3 and Grade 6 , and Panel C shows estimates for new students only.All regressions include set of individual controls (own test score in Grade 3, gender, Enlglish as Second Language learner, Canadian born, and whether student learned English at home), school controls - urban school, Catholic school board, school from Toronto Metropolitan Area; all regressions include neighborhood controls - log of median household income, proportion of residents with university degree, some university, proportion of low income families, and proportion of recent immigrants. Regression in Panel B also controls for the average achievement of "old peers".

Table 6. Summary Statistics (Apparently Random Assignment)

	More than one class	One Class
Math test score, Grade 3	2.8 (0.68)	2.75 (0.68)
Math test score, Grade 6	2.74 (0.77)	2.78 (0.76)
Female	0.50 (0.50)	0.50 (0.50)
English as a second language (ESL)	0.013 (0.12)	0.016 (0.13)
Born outside Canada	0.095 (0.30)	0.035 (0.18)
Did not learn English at home	0.2 (0.40)	0.11 (0.31)
Urban area	0.9 (0.30)	0.62 (0.49)
New peers	0.38 (0.48)	0.23 (0.42)
Number of observations	309,821	43,680
Number of schools	2,543	852

Note:The sample in the first column comprises all students in schools with more than one Grade 6 class and for whom the data on test scores in Grade 3 and Grade 6 are available. Column (2) includes only students from schools with one Grade 6 class.

Table 7. Apparently Random Assignment and Peer Effects

Dependent variable	Mathematics test score in Grade 6										
	(1) All	(2) All	(3) All	(4) Stayers	(5) Switchers	(6) Boys	(7) Girls	(8) Level 1	(9) Level 2	(10) Level 3	(11) Level 4
Average Grade 3 mathematics test score of new peers in class	0.012	0.153***	0.155***	0.285*	0.190	0.162***	0.168***	-0.074	0.072*	0.167***	0.134
	(0.007)	(0.019)	(0.019)	(0.160)	(0.225)	(0028)	(0.027)	(0.135)	(0.040)	(0.032)	(0.096)
Individual controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
School fixed effects	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year fixed effects	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
School-year fixed effects	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
FSA controls	No	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Number of observations	43680	43680	43680	33429	10251	29911	21769	1408	11827	26285	4160
R^2	0.239	0.456	0.456	0.476	0.532	0.482	0.507	0.345	0.290	0.328	0.542

Note: Standard errors clustered at school level. *** p-value<0.01, ** p-value <0.05, * p-value<0.10. In columns (1) through (3) sample includes all students in schools with only one Grade 6 class for whom data on test score in Grade 3 and Grade 6 are available. Columns (4) and (5) present estimates for incumbent students (those who did not switch school betwenn Grades 3 and 6) and switchers (those students who switch school after Grade 3) respectively. Separate regressions for boys and girls samples are presented in Columns (6) and (7). Columns (8)-(11) presents estimates for students at different levels of achievement. All regressions include individual controls (own test score in Grade 3, gender, Enlglish as Second Language learner, Canadian born, and whether student learned English at home), school controls - urban school, Catholic school board, school from Toronto Metropolitan Area; all regressions exept in column (1) include neighborhood controls - log of median household income, fraction of residents with university degree, some university, fraction of low income families, and proportion of recent immigrants in the neighborhood.

Table A1. Effect of the Average Quality of New Peers in Class on Test Scores (Reading)

Dependent variable	Reading test score in Grade 6									
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
	All	All	Switchers	Stayers	Boys	Girls	Level 1	Level 2	Level 3	Level 4
Average Grade 3 reading test score of new peers in	0.098***	0.099***	0.273***	0.051	0.104***	0.093***	0.164***	0.117***	0.077***	0.065***
	(0.002)	(0.002)	(0.012)	(0.005)	(0.006)	(0.006)	(0.014)	(0.008)	(0.006)	(0.013)
Individual controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
FSA controls	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year fixed effects	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
School fixed effects	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Number of observations	341797	341797	124685	217112	171762	170035	26513	92732	199455	23097
R^2	0.296	0.305	0.331	0.291	0.292	0.301	0.039	0.028	0.034	0.04

Note: Standard errors clustered at school level. *** p-value<0.01, ** p-value <0.05, * p-value<0.10. In columns (1) and (2) sample includes all students for whom data on test score in Grade 3 and Grade 6 are available in the data. Column (3) only includes students who switched schools between grade 3 and Grade 6. Column (4) include students who did not switch schools over the observed period. Columns (7)-(10) presents estimates for students at different levels of achievement. All regressions include individual controls (own test score in Grade 3, gender, Enlglish as Second Language learner, Canadian born, and whether student learned English at home), school controls - urban school, Catholic school board, school from Toronto Metropolitan Area; all regressions exept in column (1) include neighborhood controls - log of median household income, proportion of residents with university degree, some university, proportion of low income families, and proportion of recent immigrants. Regression in column (3) also include control for the average achievement of "old peers". Results with and without inclusion of this controls are quantitatively the same.

Table A2. Effect of the Average Quality of New Peers in Class on Test Scores (Writing)

Dependent variable	Writing test score in Grade 6									
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
	All	All	Switchers	Stayers	Boys	Girls	Level 1	Level 2	Level 3	Level 4
Average Grade 3 writing test score of new peers in class	0.128***	0.117***	0.316***	0.066	0.121***	0.114***	0.174***	0.148***	0.098***	0.093***
	(0.005)	(0.006)	(0.012)	(0.005)	(0.007)	(0.007)	(0.033)	(0.008)	(0.007)	(0.015)
Individual controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
FSA controls	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year fixed effects	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
School fixed effects	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Number of observations	343619	342837	125128	217709	172317	170520	3497	106699	209523	23118
R^2	0.246	0.255	0.279	0.246	0.292	0.301	0.051	0.051	0.061	0.047

Note: Standard errors clustered at school level. *** p-value<0.01, ** p-value <0.05, * p-value<0.10. In columns (1) and (2) sample includes all students for whom data on test score in Grade 3 and Grade 6 are available in the data. Column (3) only includes students who switched schools between grade 3 and Grade 6. Column (4) include students who did not switch schools over the observed period. Columns (7)-(10) presents estimates for students at different levels of achievement. All regressions include individual controls (own test score in Grade 3, gender, Enlglish as Second Language learner, Canadian born, and whether student learned English at home), school controls - urban school, Catholic school board, school from Toronto Metropolitan Area; all regressions exept in column (1) include neighborhood controls - log of median household income, proportion of residents with university degree, some university, proportion of low income families, and proportion of recent immigrants. Regression in column (3) also include control for the average achievement of "old peers". Results with and without inclusion of this controls are quantitatively the same.

Table A3. Effect of the Average Quality of New Peers in Class on Test Scores - Various Samples

Dependent variable	Mathematics test score in Grade 6					
	(1)	(2)	(3)	(4)	(5)	(6)
	All classes	More than 1 Grade 6 class in school	Classes with no new peers in Grade 6	Classes with only new peers in Grade 6	Classes with more than 10 students	Classes with less than 10 students
Average Grade 3 math test score of new peers in class	0.133***	0.134***	0.107***	0.400***	0.095***	0.236***
	(0.008)	(0.007)	(0.032)	(0.022)	(0.008)	(0.014)
Individual controls	Yes	Yes	Yes	Yes	Yes	Yes
FSA controls	Yes	Yes	Yes	Yes	Yes	Yes
Year fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
School fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Number of observations	353227	309547	8585	57912	331058	22169
R^2	0.3	0.32	0.32	0.36	0.31	0.36

Note: Standard errors clustered at school level. *** p-value<0.01, ** p-value <0.05, * p-value<0.10. Column (1) is a baseline result of fixed-effect estimation and presented here for comaprison. Sample in column (2) includes all school with more than 1 Grade 6 class in year t. Column (3) is a sample restricted to classes where all students stayed in the same school since Grade 3. The mean peer achievement is an average test score in Grade 3 of all students except student i. Column (4) sample includes classes with new students only. Columns (5) and (6) stratify sample by the number of students in class - with more than 10 (column (5)) and less than 10 students (column (6)). All regressions include individual controls (own test score in Grade 3, gender, Enlglish as Second Language learner, Canadian born, and whether student learned English at home), school controls - urban school, Catholic school board, school from Toronto Metropolitan Area; all regressions include neighborhood controls - log of median household income, proportion of residents with university degree, some university, proportion of low income families, and proportion of recent immigrants. Results with and without inclusion of this controls are quantitatively the same.

Appendix

This appendix describes the process of allocation of students into classrooms using information I have collected through interviews with school principals and in online survey of elementary school principals. The sample size of the survey is very small; still, it provides useful insights into the general allocation rules of assigning new students to classrooms. The schools whose principals I interviewed are located in Toronto and belong to the Toronto District School Board. Below, I present an online survey with break-down of responses as well as Principals' comments.

1. Some schools organize instruction differently for students with different abilities. What is your school policy about this for students in Grades 3 to 6? Please check all that applies.

   (a) Students are grouped by abilities for all subjects (0.0%)

   (b) Students are grouped by abilities for some subjects (64.3%)

   (c) Students are NOT grouped by abilities (35.7%)

   (d) Open-ended answer:

      - We group children for targeted instruction in some areas - eg language or math, where they may need a more intensive level of support or review or direct instruction around some concept or skill in which there is a gap.
      - mixed ability groupings are what most teachers strive to achieve unless focussing on a specific strategy.
      - we may group students by abilities for skills or reading levels from time to time
      - This is determined by the top two categories and many other factors in the goal to set them up for success
      - Teachers differentiate their instruction and group based on that information.

- Students are grouped by abilities for some subjects for in some situations
- Consideration is given to abilities - academic & self-regulatory, personalities, gender, group dynamics, etc., with a strong preference for heterogenous and balanced classes where every teacher shares in the joys and challenges of various student aptitudes and needs and every class has strong peer role-modelling for every area.

2. What is your school policy, if any, for assigning NEW students to classes? Please check all that applies.

    (a) New students are assigned to classes based on their prior achievement (report card) (11.1%)

    (b) New students are assigned to classes based on the class size capacity (96.3%)

    (c) New students are randomly assigned to classes (3.7%)

    (d) There is no specific policy for assigning new students to classes (3.7%)

    (e) Open-ended answer:

    - We also take into consideration the needs of the incoming student, the composition of the existing class and the schedule.
    - Needs of the student
    - Unless they are an exceptional student (identified).
    - We generally do not have the prior report card until after the student has arrived at the school so prior achievement can't be taken into consideration.
    - based on a number of factors, including those above
    - We also look at any Special Education or ESL designations
    - Determined usually by class size
    - Students are placed based on many factors, some listed above, all with the goal for setting them up for success

- However, needs are taken into consideration to support each child;s learning
- very small school - only one of each class
- New students are assigned to classes based on a number of factors.
- I consider all information available and try to maintain balanced #s and balanced classes.
- students with ESL needs are also placed with students who speak the same language as them
- Students are placed to create heterogeneous groupings

3. The growing body of quantitative evidence shows that having more female students in class positively affects academic achievement of all students. Do you have a school policy about this?

   (a) Yes, when assigning students to classes, male to female ratio of students in class is taken into account (40.0%)

   (b) No, when assigning students to classes, male to female ratio of students in class is NOT taken into account (16.7%)

   (c) There is no specific school policy but students are assigned to classes to preserve approximate 1:1 ratio of male to female students (43.3%)

   (d) Open-ended answer:

   - No, we do try to balance the M and F ratio in each class at a specific grade level. This is almost never a 1:1 ratio.
   - Children do not come in evenly gendered numbers. We have experimented with some single gender classes, but generally try to balance male and female numbers in regular classrooms.
   - a M-F balance among all classes in a given grade level is strived for
   - Sometimes it is difficult to have more females than males due to numbers but an important consideration

- Our goal is to have a balance of males and females
- Do our best to balance gender, learning styles, interests and abilities