

Peer Effects in the Classroom: Evidence from New Peers

Margarita Pivovarova*

This draft: August 30, 2013

Abstract

Peer effects in education are of interest to parents, policy-makers, and researchers alike. However, there are serious obstacles to estimating these effects, notably selection, endogeneity and reflection problems. In this paper I assemble a rich longitudinal data set of individual-level test scores and introduce an innovative research design to circumvent these identification problems, and estimate the extent and structure of ability spillovers among elementary students in the same classroom. My identification strategy is based on isolating the impact of new students to a school. These new students can plausibly be viewed as randomly assigned to a class *within* their new school. I implement this strategy by using the new students' lagged test score (from Grade 3 in their previous school) as an instrument for the average performance of Grade 6 students in their new class. I simultaneously account for the potentially endogenous selection of students into schools, and school-specific time trends, by incorporating school-by-year fixed effects. Using administrative test scores data for three cohorts of Ontario elementary students (observed in Grades 3 and 6 for three subjects - mathematics, reading and writing), I find positive and statistically significant peer effects - a one standard deviation in the classmates' average test score leads to 0.25 standard deviation increase in individual test score. However, the effect is not constant: the impact of classmates' average ability varies with a student's individual ability. I also contrast the effect of class level peers with the impact of school level peers and find that it is immediate classmates as opposed to school or grade level peers who matter for individual achievement in elementary school. This finding suggests that the definition of the relevant peer group is important for recovering peer effects in education.

JEL CLASSIFICATION: J10; J15; I20

KEYWORDS: peer effects, education.

*Assistant Professor, Arizona State University. e-mail: margarita.pivovarova@asu.edu. I am indebted to my supervisor, Dwayne Benjamin, for inspiration and continuous support during this project, and to Gustavo Bobonis and Philip Oreopoulos who provided valuable suggestions that significantly improved the original draft. I benefited from insightful comments from Josh Lewis, Elaine Liu and seminar participants at the University of Toronto. I thank the Educational Quality and Accountability Office and especially Michael Kozlow, for providing the data used in this research. Financial support from the Royal Bank Graduate Fellowship in Public and Economic Policy is gratefully acknowledged.

1 Introduction

*There's a new kid in town
Just another new kid in town
Everybody's talking 'bout the new kid in town
Everybody's walking like the new kid in town*

(The Eagles, 1976)

School administrators recognize the importance of interaction between classmates for individual outcomes. That is why schools have various rules for assigning students to classes – some group them by abilities (tracking), while others balance based on students' ability and other characteristics. What classroom arrangement is the most efficient for raising the overall achievement in a classroom? While the answer is not clear, most of the debates about school integration or segregation are based on the premise that schoolmates have an impact on individual academic and behavioral outcomes. In this paper, I exploit the institutional features of public education in Ontario, and combine a rich and unique data set on students' test scores and innovative research design to estimate the impact of classmates on individual achievement. I find positive and sizable spillovers among students in a classroom. I also provide evidence about the structure of ability spillovers in elementary school and show that peer effects are monotone and increasing in own achievement level.

Peer effects in education received a great deal of attention because of their importance for parents and policy-makers.¹ An increasing number of recent studies suggests the existence of ability spillovers in education. The key challenge of these studies is the selection of students into peer groups and endogeneity of the group outcome: the composition of schoolmates may reflect the sorting of households across neighbourhoods. If parents choose neighborhoods with better schools, there will be a positive correlation between a student and her classmates achievement that could be misinterpreted as a peer effect. Even when selection into schools can be controlled for, there may still be non-random sorting of students into classes *within* a school. If students are grouped in classes by abilities, the positive correlation between

¹See Sacerdote (2011) for the most recent review of peer effects in education literature.

achievement level of classmates could be mistaken for peer effects even when there are no ability spillovers. The opposite may happen if students are assigned to balanced classrooms. This will generate a negative association between student’s own achievement and the average achievement of classmates. In this case, when peer effects exist, they may be hidden by the non-random “negative” sorting of students into classes.

The problem of selection into schools and classrooms potentially can be solved by randomly assigning students to schools and classes and measuring their pre- and post outcomes. Examples of such experiments exist at a college level (Sacerdote, 2001; Zimmerman, 2003; Carrell *et al*, 2008), but are rare at the elementary or secondary school level.² When the random assignment is not possible, researchers turn to the variation in peer characteristics across adjacent grades or cohorts within the same school.³ These studies focus on the impact of background characteristics of schoolmates on educational outcomes, as opposed to ability spillovers. Due to data constraints, the peer group in these studies is identified at a grade or cohort level, as opposed to classroom level where the majority of the interactions between elementary students takes place. Classroom-level studies of peer effects are rare because it is often not feasible to match students to classrooms, and because students may still be sorted into classes *within* a school. School administrators “mix and match” students with different abilities and background characteristics to achieve efficiency and better aggregate outcomes. Parents also may influence the placement into classes if they think that a specific teacher or a classroom would provide more learning benefits to their children. To overcome this latter selection problem, I exploit the natural turnover of students in schools as a source of plausibly random change in the composition of the classmates.

To provide intuition for the identification strategy in this paper, consider a hypothetical

²A notable exclusion is Tennessee’s Student Teacher Achievement Ratio project (also known as Project STAR), a large scale randomized experiment carried out in 79 schools in Tennessee when elementary students were randomly assigned to classroom of various sizes, or large classes with additional full-time teacher aide. Boozer and Cacciola (2001) and Whitmore (2005) exploited the random assignment of students participated in Project STAR to estimate peer effects in elementary education.

³Hoxby, 2000; Lavy and Schlosser, 2011 exploited variation in gender composition of schoolmates; Angrist and Lang, 2004 and Hanushek et al, 2009 - variation in race; Friesen and Krauth, 2011 - variation in home language.

school with two identical classes. Two new students enter the school at the start of the academic year. The new students are different in their abilities and achievement - one is an A+ student while the other struggles just to pass. The school principal who makes a decision about the allocation of students into classes, flips a coin to determine which class would get which student. By a random draw, one class gets the “bad” student and the other class gets the “good” student. Nothing else changes in the classroom environment except for a new classmate who shifts the average quality of peers for everyone else. At the end of the year, all students in school take a standardized test. Since the new students were assigned to classes by a random draw and not based on class or teacher characteristics, and everything else was held constant, the difference in the test score of students in two classes relative to the baseline can be attributed to peer effects.⁴

The identification strategy in this paper simulates the hypothetical situation above. In order to do that, I argue and empirically demonstrate that the initial assignment of *new* students to classes *within* a school can be considered random conditional on observed background characteristics of both new students and students who stayed in the school (incumbent students). For instance, the data confirm that new students are allocated to classes primarily based on their gender and first language. New girls are more likely to be allocated to a class where there are too many boys, and students whose first language is not English are “bundled” together. There is no evidence however, that new students are allocated to classes based on their lagged achievement or lagged achievement of other students in that classroom. In fact, the quantitative findings in section 3 show that there is no relationship between the past test scores of the new and incumbent students. This is in the contrast with the evidence that the contemporaneous test score of the incumbent students is highly correlated with the lagged achievement of the new students in a class. I then compare test

⁴The recent theoretical advancements in peer effects literature suggest that the impact of classmates might be different depending of what are the underlying mechanisms of those effects. For instance, Liu, Patacchini and Zenou (2011) propose that peer effects might stem from two different sources: the total value of the quality of the peer group or aggregate average; and another one which results when deviation from the norm in the group, or the *average* quality of the group, is costly. However, the mechanisms of the peer effects are outside the scope of this paper.

scores for students in classes with different distribution of new students within the same school taking into account selection into schools and neighborhoods.

This strategy helps me to solve the two key problems - non-random selection of students into classes within a school and simultaneity of achievement among classmates.⁵ First, if new students are allocated to classes based on their observed background characteristics and not on the lagged achievement, then the change in the ability distribution of classmates can be regarded as good as random. The new students in class thus generate a non-systematic variation in the average achievement levels between classes within a school. This variation can be used to estimate the impact of classmates on individual achievement. Second, the lagged achievement level of new classmates does not “reflect” the contemporaneous achievement of students in a classroom and is immune to the reflection problem since it had been realized before these students entered their new school.

To estimate peer effects, I use observational data on test scores of students in elementary school. One of the main critiques of using observational data for study of ability spillovers in schools is that it is not immediately clear whether and how the findings from these data can be used to inform policy decisions.⁶ Compared to observational data, random assignment of students to classes could inform what policies might work to raise the achievement in a classroom or in school. However, the generalizability of such experiments is questionable.⁷ Conversely, observational data, especially when it cover the entire population, help to understand what classroom arrangement does work and how does it work. The aim of this paper is not to suggest policy interventions, but to analyse what classroom arrangements work in elementary school, and who benefits and who loses from such arrangement.

⁵The simultaneity is also known as “reflection” problem and describes the reciprocal nature of the outcomes of group members. The problem is that it is not possible to infer the direction of the influence in the observed outcomes - whether it is an individual who influences the group, or it is a group average that influences the individual.

⁶See Carrell, Sacerdote and West (2012) who show the danger of deriving policy prescriptions from the reduced form estimates.

⁷While not directly related to this study, Deaton (2010) in the context of development argues that experimental evidence with some exceptions have no statistical superiority and produce knowledge that are too narrow and too local to derive policy implications.

I make three contributions to the peer effects literature. First, I suggest a new approach to overcome selection and endogeneity in peer effects' estimates. I apply this approach to a rich panel dataset for the entire population of elementary students in Ontario, the largest Canadian province. The data contain the unique information about students' transition to schools and complete records of test scores on provincial standardized tests. The data also allow to track students between two waves of assessment and match them to classrooms. This study is the first to use these data to estimate peer effects in elementary education.⁸ I estimate positive economically and statistically significant ability spillovers among students in a classroom within elementary school. I show that one standard deviation increase in the average achievement of classmates results in a 0.25 standard deviation increase in a student's test score. When I account for the endogenous sorting of students into classrooms, I find that the OLS estimates are downward biased because of the balancing students with different abilities in a classroom.

Second, I compare estimates of the impact from grade-level peers to classroom-level peers. I do not find significant spillovers when I aggregate peer group to a grade level. This finding points out that the definition of the relevant peer group is important in recovering the estimates of peer effects. It also helps to reconcile the mixed findings from other studies that documented small or no effect from peers at a school level in elementary schools and studies that found larger effects using classroom as a unit of analysis.

Third, I explore the structure of peer effects and find that the impact of classmates is not constant but varies with individual achievement. For instance, the largest effect from the average achievement of classmates is observed for students who are high-achievers themselves. The distribution of abilities in a classroom also has an impact on individual achievement, but the effect is small and becomes not significant when the average achievement in a class is taken into account. Consistent with previous studies, I also find that girls benefit more from good peers than boys.

⁸The same data have been used in a study of school competition by Card, Dooley and Payne (2010) and in a research of the impact of teachers' strikes on student achievement by Baker (2012).

The rest of the paper is organized as follows. In the next section, I lay out the identification strategy and explain how I use the changes in the composition of classmates to identify peer effects in the classroom. I show how this strategy helps to overcome non-random sorting of students into classes and simultaneity of peers' outcomes within a school. Section 3 describes the institutional background and the nature of test scores used in the study. In that section, I also explain what unique features of the data help to identify peer effects. I then demonstrate that the assignment of new students into classrooms can be considered random conditional on observed characteristics, and show that the quantitative findings from the data are aligned with the information collected from the survey and interviews with school administrators. Section 4 starts with the description of main results. I apply the instrumental variables strategy and show how it helps to overcome the endogeneity bias in the OLS estimates. I then test whether peer effects are heterogeneous and show that the impact of classmates varies with individual achievement. In section 4 I contrast the estimates of peer effects obtained when the relevant peer group is defined at a classroom and at a school or grade level, and provide the rationale for why these estimates might differ. As shown in section 4, the main results of the estimation are robust to the definitions of new classmates, sample selection, and the outcome measure. Section 5 summarizes the findings and discusses directions for future work.

2 Identification Strategy

I start by laying out the basis for the empirical specification. I then explain the benefits of using instrumental variables strategy (IV) to identify ability spillovers among students in a classroom. The model most widely used in the social interactions literature is a structural model from Manski (1993):

$$Y_{ij} = \alpha X_i + \beta \bar{X}_{(-i)j} + \gamma \bar{Y}_{(-i)j} + \theta Z_j + \varepsilon_{ij} \quad (1)$$

where Y_{ij} denotes outcome of individual i from group j , X_i is a vector of background characteristics of individual i , \bar{Y}_j , \bar{X}_j and Z_j are the group mean outcome, average group characteristics and common group factors respectively. In the academic achievement literature, coefficient β captures the impact of the exogenous characteristics of a student's classmates on his/her test score and is referred to as contextual or exogenous effects parameter. Coefficient γ measures the effect of the average group achievement on individual's outcome and is known as endogenous effect. θ is a correlated effect parameter and represents the impact of the common group factors - neighbourhood, school, class.^{9 10}

Since parameters β and γ of the structural model in (1) cannot be identified separately, the standard way to approach the problem is to estimate a reduced form of the structural equation and assume that either exogenous or endogenous effect is not present. The basic estimating equation that describes the impact of the average quality of classmates on student i 's test score is a reduced form of the structural model in (1):

$$Y_{icst} = \alpha_1 \bar{X}_{(-i)cs} + \alpha_2 X_i + S_{st} + \varepsilon_{icst} \quad (2)$$

where Y_{icst} is an outcome of interest of student i at time t , \bar{X}_{-icst} is the average characteristics of all students in the same class except student i 's contribution, and X_i and S_{st} are vectors of student i 's background characteristics and school characteristics respectively. The error term, ε_{icst} can be decomposed into the individual unobserved heterogeneity, v_{icst} , classroom level idiosyncratic error, ζ_c , common shocks to a group at a school level, ν_s . One

⁹Researchers are often interested in identifying exogenous and endogenous parameters of the model, and therefore in some of the studies, the third group term, Z is often omitted. One notable exclusion is Bobonis and Finan (2009) who study peer effects in school attendance in Mexican villages under the random assignment into conditional cash transfer program. The absence of Z in equation (1) does not change the interpretation of the identification problems with the empirical model based on (1).

¹⁰In the majority of the empirical work on peer effects, the interest is in the reduced form of the structural model in equation (1). However, the coefficients β and γ can not be separately identified unless the specific assumptions are imposed. The reduced form is given by $Y_{ij} = \alpha X_i + \left(\frac{\beta + \alpha\gamma}{1 - \gamma}\right) \bar{X}_j + \left(\frac{\theta}{1 - \gamma}\right) Z_j + \varepsilon_{ij}$. I direct the interested reader to the original derivation of this reduced form equation to Manski (1993) and to the detailed discussion of the identification problems with the linear-in-means model in Blume, Brock, Durlauf and Ioannides (2010).

can also add combinations of school, classroom and year interactions to the model. In Manski’s original terminology, α_1 captures either exogenous or endogenous effects which cannot be separated without imposing additional restrictions. The term S_{st} represents school correlated effects. To credibly identify the causal effect of classmates on student i ’s outcome, the average characteristics of classmates, \bar{X}_{-icst} should not be correlated with the unobserved determinants of individual achievement captured by the error term in equation (2).

The ideal experiment to identify the effects of classmates on individual outcome would be to randomly assign various $\bar{X}_{(-i)cs}$, or classmates, to “identical” students and measure the achievement level before and after the experimental period and compare the outcomes of identical students who were exposed to different classmates. However, the large-scale experiments with random assignment of students into classrooms are rare¹¹ and have their own limitations.¹² In the absence of random assignment, one needs to find a source of plausibly exogenous variation in the average quality of classmates to estimate the impact on student i ’s outcome with equation (2).

Without exogenous variation in the average quality of classmates, the estimate of peer effect in (2) would be biased for at least three reasons. First, if there are unobserved determinants of individual achievement that are correlated with the average characteristic of the peer group, $\bar{X}_{(-i)cs}$, then the impact of peers is likely to be under- or overstated.¹³ For instance, parents who care about the education of their children could be consistently enrolling their children to a class with the presumably most effective teacher. The coefficient

¹¹The studies of peer effects in education that exploit the random assignment of group members are Sacerdote (2001) for Dartmouth College roommates and Carrell, Fullerton and West (2008) for the US Air Force Academy students randomly assigned to squadrons. The purely randomized experiment was designed by Duflo, Dupas and Kremer (2011) in Kenya when students were assigned to either tracked or mixed classes.

¹²While the random assignment remains the golden standard of the identification, Arcidiacono *et al* (2005) for instance, show that peer effect estimates may be biased downward among randomly assigned classmates, and that some degree of sorting might enable more precise estimates of peer spillovers. Also, from the practical point of view, large scale experiments in elementary schools are not feasible.

¹³This 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).

on peer variable in this case would simply document that children whose parents care about education end up in the same classroom, but not identify the impact students have on each other. Under such scenario, we are likely to “find” peer effects even if there are no ability spillovers among students. One way to account for the unobserved determinants of individual achievement is to include a lagged test score as an additional control. If the lagged test score is a good predictor of contemporaneous achievement, then it captures the unobserved family inputs, child’s ability, motivation and effort.¹⁴

Second, the contemporaneous average test score of student i ’s classmates is endogenous as it is likely to be determined together with student i ’s own outcome. In the peer effects literature, there are two popular solutions to account for endogeneity of the peer group variable in equation (2). The first one is to use a lagged outcome of group members without student i ’s contribution, $\bar{Y}_{-ics,t-1}$.¹⁵ In this case, the lagged outcome of classmates is assumed to be independent of student i ’s current outcome. This strategy is only valid if the lagged group outcome was realised when student i was not a member of that group or the group has been formed randomly. This assumption is often violated in cross-sectional data: the lagged achievement of classmates is likely to be correlated with the contemporaneous individual achievement because the current classmates of student i were his/her classmates in the previous period.¹⁶ Even with panel data, it is often not possible to track classmates from year to year.

The second strategy to address the endogeneity of the average peer group outcome in equation (2) is to use the plausibly exogenous variation in the background characteristics of

¹⁴Later in the empirical analysis I will show that prior achievement is the most important determinant of the current test score which solely explains 27% of the variation in the contemporaneous test score. Adding other relevant controls and school and year fixed effects brings the explanatory power of the regression to a maximum of 46%.

¹⁵This strategy has been used in Hanushek *et al* 2003, Betts and Zau 2004, Vigdor and Nechyba 2007, Burke and Sass 2011.

¹⁶As Gibbons and Telhaj (2012) point out, in a standard educational value-added specification which controls for individual prior achievement and uses average lagged outcome of current classmates to identify peer effect, the estimate of α_2 will be downward biased if the outcomes of the individual and some of the classmates have been jointly determined by the same factors.

adjacent cohorts of students within a given school, $\bar{X}_{(-i)st}$.¹⁷ This strategy has been widely applied for the identification of peer effects in elementary and high school.¹⁸ This strategy assumes that changes in gender, race, or other background characteristics of schoolmates across adjacent grades or cohorts within a given school are purely random and are orthogonal to unobserved determinants of individual achievement. In other words, studies based on this strategy compare cohorts with different fractions of girls, boys, problem students, etc., and estimate the impact of the average quality of students in the same grade and in the same school on student i 's outcome. The spillovers estimated using this approach stem from the average background characteristic of classmates, such as gender or race, and not from the actual measures of achievement, and are usually referred to as exogenous effects.¹⁹ Studies based on cohort-grade variation in peers' background assume by design that schoolmates are the relevant peer group and do not estimate peer effects at a finer level - classroom.

The third identification issue in equation (2) is a non-random sorting of students into schools that introduces selection bias into the OLS estimate of peer effect. The logical step would be to include fixed effects at a level higher than the level of the relevant peer group to account for non-random sorting.²⁰ For instance, in order to estimate classroom level peer effects, one needs to include school fixed effects, otherwise peer effects are absorbed and cannot be identified (Ammermueller and Pischke 2009). This approach solves selection into peer groups when the relevant peer group is identified at a school or grade-cohort level, but not at a classroom level. The sorting of students into peer groups or classes within a school remains an issue. The selection of students into classes within a school can be dealt with by

¹⁷Hoxby (2000) popularized the cohort-to-cohort variation strategy and the number of studies followed using the same methodology.

¹⁸See Angrist and Lang, 2004; Hanushek *et al*, 2009; Friesen and Krauth 2011, Lavy and Schlosser, 2011 among others.

¹⁹While interesting on their own, exogenous peer effects have only limited implications for policy interventions. The main interest of policy-makers is in the behavioral, or endogenous effects, which results from the peer outcomes and can be manipulated. Behavioral spillovers also assume the existence of social multiplier.

²⁰The method has been applied in both cross-sectional and panel data studies. See, for instance, Arcidiacono and Nicholson (2005), Hoxby and Weingarth (2006), Vigdor and Nechyba 2007 and Lavy, Passerman and Schlosser (2012) among others.

either demonstrating that the assignment of students into classes is random²¹, or by finding an exogenous shifter that affects the composition of the classmates.

A different approach to estimate the strength of peer interactions in a classroom was suggested and implemented by Graham (2008). He used the experimental data from Tennessee class size reduction project STAR and compared variance in academic quality of peers across small and large classes and found large peer effects. In my study, I am not able to rely on this method as there is small variation in class size both within and between schools.

In this paper, I employ the instrumental variables (IV) strategy to estimate peer effects in the classroom. The “shifting” variable in my approach is the average lagged test score of incoming students which I use as an instrument for the current average achievement of student i ’s classmates.²² My identification strategy incorporates school, cohort and school-cohort fixed effects and simultaneously accounts for endogeneity of the peer variable using a plausibly exogenous variation in the average quality of new classmates. In particular, I use quantitative evidence that new students are *initially* allocated to classes based on their observed background characteristics - gender, English as a second language, and country of birth or first language spoken at home, and not on abilities as measured by their lagged test score.²³ The key peer variable, $\bar{Y}_{(-i)_{cst}}$ is computed as the average Grade 6 test score of all students in class except student i . The instrument, $\bar{W}_{cs,t-1}$ is constructed as the average Grade 3 test score of student i ’s current classmates who were not his/her class- or

²¹See Vigdor and Nechyba (2007) and Ammermueller and Pischke (2009) for two examples of dealing with selection at a classroom level

²²My study is not the first one to use instrumental variables strategy to estimate peer effects. For instance, Evans, Oates and Schwab (1992) use an aggregate measure of economic disadvantage to instrument for the corresponding school level. Cooley (2010a) exploits the policy of holding back students who scored below a certain level on a reading test and uses the fraction of those students in a grade as an instrument. Gould, Lavy and Paserman (2009) used predicted fraction of immigrant students in a grade to instrument the actual concentration in order to estimate the impact of having immigrant classmates on long-term outcomes. Most recently, Imberman, Kugler and Sacerdote (2012) used the achievement weighted fraction of hurricane evacuees to instrument for the average grade achievement.

²³Several papers exploited quasi-random assignment of students to classes based on country-specific policies (Kang (2007) for South Korea, Wang (2010) for Malaysia), group re-assignment (Angrist and Lang (2004), Hoxby and Weingarh (2006))

schoolmates in Grade 3.²⁴ I estimate a two-stage model of the structural equation in (1):

$$\bar{Y}_{(-i)cst} = \gamma_1 \bar{W}_{cs,t-1} + \gamma_2 \bar{X}_{(-i)cst} + \theta_s + \delta_t + \psi_{st} + v_{icst} \quad (3)$$

$$Y_{icst} = \beta_1 \bar{Y}_{(-i)cst} + \beta_2 Y_{ics,t-1} + \beta_3 X_i + \beta_4 \bar{X}_{(-i)cst} + \bar{Y}_{(-i)cs,t-1} + \rho_s + \lambda_t + \phi_{st} + \varepsilon_{icst} \quad (4)$$

where i denotes individual student, c denotes a classroom, s indicates school, and t is the year when a student took Grade 6 test, or a student's cohort.

Equation (3) implies that the average Grade 6 test score of all classmates is a function of the average lagged achievement level of new students in class along with the average characteristics of students in that classroom, \bar{X}_{cst} ; school, year and school-by-year fixed effects. Equation (4) describes the relations between the contemporaneous achievement level of student i and the average test score of all classmates in Grade 6, $\bar{Y}_{(-i)cst}$, plus included instruments from equation (3): average background characteristics of student i 's classmates, school, cohort and school-by-cohort fixed effects. Specification in equation (4) also conditions on the own lagged achievement of student i measures by the results of Grade 3 test, as well as on the average baseline achievement in the classroom, $\bar{Y}_{(-i)cs,t-1}$.

The two stage specification models the selection process into schools and classrooms. In order to see that, imagine that the first stage is when parents and students are sorted across neighborhoods and schools. This selection is taken care of by the school fixed effects at both stages of the estimation. The second stage occurs when a school principal makes a decision about the placement of new students into classes. The second stage is not a random process because the school principal allocates students based on observed child's and parents' characteristics. The second-stage sorting cannot be addressed by school fixed effects, nor can it be addressed by inclusion of classroom fixed effects as it would absorb all useful variation within a school. Given that in the data I observe the lagged achievement

²⁴In the next section I explain the construction of the instrument in detail and show which unique features of the data I use to construct the instrument, and how students who were not schoolmates before are identified in the data.

and background characteristics of new and incumbent students in a school and class, I can analyse the assignment of new students into classes. As will be evident from the data analysis in section 3, the placement of new students is not based on their lagged achievement and the average achievement of incumbent classmates. Instead, the placement is based on observed background characteristics and class size. This observation allows me to think about the placement of new students as plausibly random conditional on observed characteristics.²⁵ I employ this observation in equation (3) to isolate the part of the variation in the average ability of all classmates which is plausibly random. In equation (4), I use this part of the variation to identify the impact of classmates on individual outcome.

The model with instrumental variables includes average background characteristics of current peers, \bar{X}_{cst} .²⁶ This allows me to disentangle the impact of the background characteristics, such as gender and first language, from the impact of abilities captured by the average test score.

My identification strategy is closest in spirit to that in Imberman, Kugler and Sacerdote (2012) who exploit the influx of hurricane evacuees in Texas and Louisiana to study the impact of new schoolmates on "native" students. The authors use exogenous variation in the fractions of evacuees at a grade level and estimate peer effects by instrumenting quality of current schoolmates with the (where possible) achievement weighted fraction of evacuees. While this study is a step forward in the identification of peer effects, it has limited external validity and implications for policy design. First of all, such a large shift in the composition of class- or schoolmates as the one caused by hurricanes Katrina and Rita is rare. The naturally occurred turnover in schools is smaller and occurs continuously. Also, the pool of incoming students is more heterogeneous in my data as students tend to come from more

²⁵This is also known as conditional independence assumption, CIA.

²⁶Cooley (2010b) shows that the interpretation of the coefficient in the achievement production function differs depending on whether both the exogenous and endogenous characteristics of the peer group are included. In a case when the endogenous characteristic is a function of background peers' characteristics, she shows that the coefficient on background characteristic has a counterintuitive sign and is attenuated to 0. In empirical analysis in this paper the focus is on the endogenous effects. I do, however, observe that peer background does not have statistically significant impact on individual achievement.

geographically diverse areas.²⁷ Secondly, the test scores are not available for all evacuees, and for the Houston data, the identification comes from the variation in the fraction of evacuees as opposed to their test scores. In this study, I observe the lagged test scores for incoming students. Moreover, new and incumbent students took the same test at the same period of their life which makes the outcomes consistent and comparable across schools and cohorts. And lastly, the nature of the data in my study allows estimation of peer effects at the level of the classroom, and not only at a grade or school level.

For the identification strategy in this paper to yield an unbiased estimate of the impact of classmates on individual outcome, two assumptions must hold. First, the average quality of incoming students must be a strong predictor of the average class achievement at the end of Grade 6. In other words, it must shift peer performance enough to generate the variation between classes. I test the predictive power of the instrument in the results section.²⁸

The second assumption requires that the variation in the quality of incoming students is uncorrelated with the unobserved determinants of the contemporaneous achievement of student i . This exclusion restriction states that the only channel through which the quality of incoming students may affect the achievement of student i is through the change in the achievement level of all classmates at the end of Grade 6. The main threat to the validity of this exclusion restriction is a non-random placement of incoming students to classrooms. For instance, new students might be assigned to classes in a systematic way based on their abilities or other characteristics unobserved to researcher. As an example, consider that new students are consistently matched to specific teachers; then the error term in equation (3) will be correlated with the average ability of incoming students through the common shocks

²⁷Imberman, Kugler and Sacerdote (2012) mention that “evacuee children came from some of the worst-performing schools in the country” and that there is no complete certainty that the result would extent to other settings with typically observed variation in peer quality. In this study, as shown in Figure 1 in Appendix the aggregate turnover of students in Ontario elementary school is consistent from year to year. The average quality of the incoming students on average remains constant over three years observed in the data and is not different from the pool of incumbent students. This means that my estimates are based on a natural process of students moving from one school to another within the province as well as outside the province and country.

²⁸Section 4 reports the first-stage relationship between the lagged average achievement of incoming students and contemporaneous achievement of all students in class.

at a classroom level. Or, a school might have a policy of tracking and matching students with similar abilities to the same classrooms. In this case, the peer group is not formed randomly and past achievement of incoming students would be positively correlated with the contemporaneous individual outcome. As an opposite example, a school might prefer to mix students with different learning outcomes. In this case, the correlation between past achievement of new students and individual outcome would be negative. The degree of sorting into classes within a given school might be different, but as long as it is systematic, it would imply biased estimate of peer influence in equation (4).²⁹

In the next section, I provide evidence that the exclusion restriction is well-grounded. I relate the characteristics of incoming students and compare them to the average characteristics of incumbent students at a classroom level. To examine the exclusion restriction further, I turn to the evidence from the field: data and information collected through responses of elementary schools' principals to a questionnaire and in personal interviews.³⁰

3 Background and Data

3.1 Elementary 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).³¹ French schools

²⁹A valid concern is that classes with a small number of students are classes with on average larger share of low-achievers, and new students are allocated into those small classes because of the capacity consideration. The robust finding in the literature is that classes with low average achievement are the smallest classes in the school. I checked the relationship between the class size and average achievement conditional on variety of other available class characteristics and found no significant relationships between average achievement and class size. Give the tight distribution of class size in Ontario data and cap on a class size in elementary and middle school, the finding of no correlation seems to be plausible.

³⁰The results of the survey of school principals and their open-ended responses are reported in Appendix.

³¹There is a large sector of private schools on Ontario that are not regulated by the Ministry of Education and considered businesses or non-profit organizations. Instructors in private schools are not required to be the members of College of Teachers of Ontario. As of 2012, there are 989 private schools listed on the

represent about 10% of elementary enrollment. School boards are required to admit all students who, or whose parents or guardians, reside in the school section.³² Elementary school includes Kindergarten to Grade 8 while secondary school comprises Grades 9 to 12.

The main outcome variable in this paper is the level of achievement assigned to a student based on the result of the standardized test. The standardized testing program in Ontario was introduced in 1996. The Education Quality and Accountability Office (EQAO) administers a number of province-wide tests in all publicly funded schools since 1998. 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 (Primary Division) and 6 (Junior Division), mathematics in Grade 9, and a literacy test in Grade 10.³³ The Grades 3 and 6 tests do not count towards a student's mark and do not affect their progress or future opportunities in school. To ensure the consistency of the tests across the province, the questions in the EQAO tests are developed to cover the full range of expectations from the Ontario curriculum, and the actual content changes from one year to the next. Teachers who mark the tests are trained using the scoring guides and samples of students' work and are required to pass a qualifying test after the training. A group of teachers is then assigned a specific set of questions to ensure consistency and reliability of results.³⁴

Each child receives an individual score on each question answered. These scores are then combined to produce a score in each of the subjects. The results of the tests show the level of proficiency achieved by an individual student relative to the Ontario curriculum. The EQAO converts the raw score on each test into a level of achievement from 0 to

Ontario Ministry of Education website, 776 of which have both elementary and secondary education, and the rest are secondary schools. Those schools enroll on average 126,000 students which corresponds to slightly less than the average enrollment per grade per year in Ontario public schools. Private schools are not required to administer provincial tests and are not a part of the sample in this paper.

³²Education 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. The two main documents required for admission into school are the child's birth certificate and a proof of residential address.

³³The EQAO tests take place during two-weeks period in late May-early June. Each school determines the exact date during that period on which students will write the test.

³⁴In Appendix, I provide an example of an actual test together with the marking guidelines.

4.³⁵ Achievement at Levels 1 and 2 indicates that a student has not yet met the provincial standards. Achievement at Level 3 is considered provincial standard, and Level 4 is awarded when a student exceeds expectations. Achievement at Level 3 and 4 suggests that a student is well prepared for work in the next grade. When a student completed the test but did not demonstrate sufficient knowledge to be assigned Level 1, a value of 0 is given.³⁶ The EQAO test scores are never bell-curved and the overall level of a child depends only on his/her own achievement on a test relative to the expectations.

Teachers and principals get a report that shows how students performed in the EQAO tests. Parents also receive test results of their child's performance. The results of the EQAO tests are sent to the school where the students wrote the test, and the school forwards them to any students who have left the school, but they are not recorded in the report card and are not available for teachers and principal in a child's new school.³⁷ The aggregate results of the EQAO test scores for each school are publicly available and serve as an indicator of school performance and provide information about the effectiveness and relative improvement of the school.

3.2 Data

The data for this study were obtained from the EQAO and through the Freedom of Information request and consist of three data sets that were linked together.³⁸ The first data

³⁵The description of levels and corresponding percentage scores are presented in the Appendix.

³⁶The discrete nature of test score presents a challenge for interpretation and comparison of the results with existing literature. In quantitative analysis I predominantly use the level system of test scores especially when I classify students as high or low achievers. Along with the four-level system, I also use standardized test score with mean zero and standard deviation equal to 1 to facilitate the comparison of my results to other studies.

³⁷The last point is important for the identification strategy in this paper. A student report card (or Ontario Student Record) is the record of a student's educational progress through schools in Ontario. Principals are required to collect and include information in the report card for each student enrolled in the school. The report card is an ongoing record and is transferred if the student transfers to another school in Ontario.

³⁸A number of Canadian studies used similar administrative data for other provinces to estimate peer effects in education: Friesen and Krauth (2010) study the impact of sorting on aboriginal students' achievement and Friesen and Krauth (2011) investigate effects of peers' home language on test scores in British Columbia, Boucher *et al* (2010) estimate peer effects in linear-in-means model for schools in Quebec.

set consists of records for all students who were enrolled in Grade 6 in 2008, 2009 and 2010 school years. For these students, I know the results of their mathematics, reading and writing tests in Grade 6, whether student was excused from writing the test, if a student is in English as Second Language program (ESL), whether student has learning difficulties, gender of a student, date when entered current school, whether student was born in Canada and learned English/French at home. For all students who also sat the Grade 3 test, I know their test score in mathematics, reading and writing.³⁹

The EQAO data have three main features which together with the institutional features of public education in Ontario allow to identify ability spillovers among students in a classroom. The first such feature is the universal coverage of all publicly-funded school in Ontario.⁴⁰ Another advantage is a consistent nature of the data on test scores which reduces the probability of measurement error in the outcome measure and peer variable.⁴¹ Schools that administer the EQAO's assessments are expected to ensure that all of their students write the test. The high participation rate in the tests provides an accurate reflection of the overall achievement level in the school and the benefit of objective assessment of learning.⁴² Lastly, the test scores data allow to match students to their respective classroom and estimate peer effects at a finer level of aggregation compared to grade or cohort peers.

The second data set is a list of schools with school name, aggregate level indicators,

³⁹The data on test scores used in this study is also a primary source for the Report Card on Ontario's Elementary School published annually by Fraser Institute. The Report Card's main goals are to help parents make a choice. The methodology used by Fraser Institute is based on ten indicators derived directly from the EQAO data - average school score in reading, writing and mathematics for Grade 3 and Grade 6, gender gap in achievement and the change in the aggregate test score. They also developed a socio-economic adjusted measure of achievement which is the difference between the predicted and observed average score based on a number of socio-economic indicators of the school neighborhood. I use the same indicators in my study to control for the neighborhood correlated effects.

⁴⁰For instance, using data from PISA and linking it to administrative records for students in England, Micklewright et al (2012) show that attenuation bias rises as peer sample size falls.

⁴¹An important issue in the studies of peer effect is a measurement error in self-reported indicators of ability and background characteristics of classmates. As Ammermueller and Pischke (2009) show, the measurement error can significantly affect the estimates of the peer effect if left not taken care of.

⁴²Indeed, the participation rate in the EQAO's assessments is high: on average, out of all students who were expected to write the EQAO's tests in Grade 6, only about four percent of students did not write or were exempted from writing. For Grade 3 tests, the fraction of students who were exempt or did not write the tests is higher by two percent on average.

and its mailing address. I matched student level data to schools based on the aggregate variables to obtain a panel of school and students records for three cohorts of six graders.⁴³ The school postal code allows to identify school neighborhood and link it to the variables from the 2006 Canadian census from the third dataset.

The third data set I use is a file which contains socio-economic indicators of school neighborhood from the 2006 Canadian census identified by the three first digits of the postal code, also known as Forward Sortation Area (FSA). These indicators 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. The characteristics of the community while linked to each school represent the average characteristics of the students' residential neighborhood and thus serve as a reasonable proxy for a student's socio-economic status.⁴⁴

I combined all three data sets into one aggregate file that contains unique information about each student, his/her classroom, 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 score data.

Additionally, in order to learn more about the placement process of students into classes within school, I conducted my online survey of school principals and also interviewed them personally. The results of the survey are described in Appendix.

⁴³Card, Dooley and Payne (2010) used a similar procedure to combine student level and school level data, but matched schools on enrolment numbers, while I matched on a vector of indicators including proportion of girls, proportion of students at each level of achievement in EQAO test, proportion of foreign born and ESL students.

⁴⁴A school postal code may not reflect a student's residential neighborhood but rather the neighborhood where a student goes to school. Given the admission requirements into elementary school, it is likely, however, that a student attends a school in the same neighborhood where she lives. Moreover, it's only secondary school that allow 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.

3.3 Sample restrictions

There is number of restrictions I need to impose on a sample to be consistent with the requirements of the empirical strategy and identification assumption. Table 1A in Appendix presents a description of samples with various restrictions and summary statistics of the main variables used in the analysis.

First, I restrict the sample to students for whom both Grade 3 and Grade 6 test scores are available. The proportion of students without a test score is roughly constant for all three cohorts and varies from 3.4 to 3.9 %. Those are students who either did not sit the test or were exempt from taking the test.

Next, I consider the enrollment of new students into schools. There are three types of schools within elementary education in Ontario: schools that enrol students in K-8, schools with K-5 only, and middle schools with Grades 6 to 8.⁴⁵ In the original sample, 41.8% of all students moved to a new school over the observed three-year period.⁴⁶ The share of students who moved to a new school in Grade 6 is 22%. About one third of these students transitioned to a middle (or feeder) school and face only new schoolmates. I drop observations from the middle schools because new students in these schools are likely to be classmates and schoolmates in the previous period which violates the main identification assumption.⁴⁷ Of the remaining 14% of new classmates, 2% moved from outside Ontario and do not have information about their lagged test score. I further restrict sample to schools with the proportion of new students per year to no more than three standard deviations of the average turnover among all schools. On average, every year a school receives an eight percent of new students. This fraction is approximately the same for each of the three

⁴⁵There are also schools with Grades 7 and 8 only, but those are rare and students from those schools are not included in the EQAO test scores data.

⁴⁶For comparison with other studies that used movers to identify peer effects, Lavy, Silva and Weinhardt (2012) report that more than 80% of students in English secondary school face new peers as a result of a natural transition process. Imberman, Kugler and Sacerdote (2012) report that the influx of hurricane evacuees in Houston and Louisiana resulted in an average fraction of evacuees at a grade and school level of 2.7%.

⁴⁷I drop from my sample 396 intermediate schools that enrolled only new students over three observed years. The total number of students enrolled in the feeder schools over three years is 39,822 or 9.3% of the entire sample.

cohorts of students. From this follows that an average incumbent six grader in my sample faces about 24% of new classmates over the period of three years.⁴⁸

While large, the fraction of new kids entering school at any given year is consistent with the data on residential mobility in Ontario. According to 2006 Canadian Census, 10.42% of households with children from 6 to 14 years old moved within the province over the last year. Of them, 7.4% moved within the same municipality or city. 36.77% of households reported that they moved within the last five years. These numbers confirm that the switching of the school by new students accompanies a residential move of their families and, thus, rule out the possibility that the turnover observed in the data is the result of only school switching within the same residential area.

I drop classes within a school with only new students, because the formation of these classes is likely to be based on school-specific rules which are not observed and may be systematic. For instance, exceptionally good teachers may be assigned to these classes to facilitate the process of children's involvement in their new school. Classes with only incumbent students are also dropped from the sample. Since my estimation strategy is based on within school variation in the peer composition, I only keep schools with more than one class in a given grade. I later use the sample of schools with only one Grade 6 class for a robustness check.

The final sample consists of 228,947 Grade 6 students in 12,556 distinct classrooms where at least one new student entered a school at the start of Grade 6. These students are enrolled into an unbalanced panel of 8,135 school-year observations.⁴⁹

The basic summary statistic for the sample of students in the data used in the estimation is reported in Table 1 and described in the following section.⁵⁰

⁴⁸I also checked the data for earlier years where I cannot link students' test score between two assessments, but still can identify new students in a classroom. The fraction of incoming students is roughly the same. In Appendix, I present a histogram of the fractions of new students in a classroom by grade and year.

⁴⁹The balanced panel of schools consists of 1,989 schools with new students in all three cohorts for a total of 5967 school-year observations.

⁵⁰Table 1 in Appendix describes all students in the data, and contrast them to the selected sample as well as to the samples with different restrictions.

3.4 Construction of peer variable

The unique features of the data in this study enable me to construct a variable which I use as an instrument for the average contemporaneous level of achievement in the classroom. This variable is an average lagged achievement of new students in a class. Before I explain the construction of the variable, I need to address two questions about the peer group: how to define the relevant peer group for student i , and how to measure the quality of the relevant peer group. The data allow me to use the smallest level of aggregation - classroom - to identify peers.⁵¹ The classroom in elementary and middle school is defined as all students who attend classes together over the school year. Unlike in high school where students are grouped for subjects and the peer group is different for each of the subjects, in elementary and middle school each student faces the same set of classmates over the year for all classes s/he is taking. Also, there is no choice of subjects in elementary and middle school for regular students.⁵² Each classroom is assigned a so-called “home room” teacher, who teaches the core subjects - mathematics, English and science. In addition, the majority of the extra-curricular activities in elementary school are organized by classroom. Thus, the common schedule, the same home room teacher, and same extracurricular activities result in the students spending six hours a day and at least 194 days per year in the same class, making the classroom a perfect setting to study peer effects.

One of the advantages of the EQAO dataset is that it also contains the date when a student entered his/her current school. Based on this information, I identify as new all students who entered a school one, two or three years before the Grade 6 test took place. I do not know a school from which a student had transferred but I know whether a child moved from another school in Ontario or from another province.

⁵¹The accuracy of what researchers identify as a set of relevant peers is important for the ability to identify peer effects. The question has been raised in Carell *et al* (2008) among others. It is rare when the peer group is explicitly observed in the data - one example of such data set is Add Health when respondents are asked to name five of their friends. In all other cases the data impose constraints on the definition of the peer group. I will come back to this question in section 4.

⁵²Students who learn English as a Second language attend special classes in addition to the regular curriculum. Also, students designated as gifted or special education have a special education plan. The results in my study are robust to inclusion of those students into the estimation sample.

For the main estimation results, I define as new only those students who entered at the start of Grade 6, but not in Grade 4 or Grade 5. The reason for this restriction is that classes, as a rule, are re-arranged at the end of each school year. Students who are classmates in Grade 6 most likely were not classmates in Grades 4 and 5. Also, new students who entered school in Grade 4 or Grade 5 can only be defined as new *schoolmates*, not classmates - the point I will later come back to. The identification strategy in this paper, which replicates the hypothetical situation described in the introduction as closely as possible, requires that new peers are randomly allocated to classrooms within a school. For students who entered school two or three years before the Grade 6 test, this requirement is less likely to hold. These students may have been systematically re-assigned to their current classrooms.

I construct two peer variables in this study. The first one is the contemporaneous average test score of all students in Grade 6 class except student i . The other one is the lagged average test score of all new students who entered school at the start of Grade 6. I use their average mathematics test score in Grade 3 for the main estimation results. I also use reading and writing test scores of incoming students for sensitivity checks and comparison with the main results.

Table 1 describes the students' population broken down by new and incumbent students and by cohort.⁵³ Of the entire sample for three cohorts of students, about 24% have entered their current school after Grade 3.⁵⁴ This fraction is constant across cohorts. There is also a relatively constant inflow of students in each grade: about 7.5% enter after Grade 3, about 8% after Grade 4, and 9% entered after Grade 5.⁵⁵ On average, the test score for all three subjects in both Grade 3 and Grade 6 is lower among new students. New students are more

⁵³Cohort corresponds to the school year when students were completing Grade 6.

⁵⁴The fraction of students who transitioned to a new school reduces by 17% compared to the original sample when I remove all schools enrolling only new students. 24% seems to be a reasonable number for a natural turnover of students in schools.

⁵⁵There is a valid concern that the relative number of new students in class is endogenous. To show that the fraction of new students in class is not related to the prior achievement of incumbent students, I run a falsification test by regressing a lagged achievement of incumbent students in a class on the share of new students in that class in Grade 6. I do not find any significant relations between the share of new students and lagged achievement of incumbent students in that class controlling for school and cohort fixed effects.

likely to be foreign-born and learn English as their second language. Given that the test score in Grade 3 for these students is recorded, it indicates that they attended school in Ontario for at least 4 years before they are observed in the data. The larger fraction of ESL learners among movers might explain why the English reading and writing test scores are on average lower among this group of students. It cannot, however, explain the statistically significant difference in the mathematics test score. Among incoming students, a smaller fraction is well prepared to work in the next grade - the share of new students at Levels 3 and 4 of achievement in mathematics is roughly 8% lower than among incumbent students.⁵⁶

3.5 Assignment of new students to classrooms

In this section, I investigate how incoming students are assigned to classes within their new school. An average school in my final sample has two or three grade 6 classes that together enroll about 50 students. Every year, a school receives 5 new students in Grade 6 and each class gets about two new students. With the average number of students in class being 21, that makes a proportion of new students equal to 9.5%.⁵⁷

Figure 1 compares the distribution of the average grade 3 mathematics test score of new students in the classroom (shaded bars) and that of students who stayed in the same school since at least grade 3 (outlined bars). The average is computed as the sum of test scores for all students in one of the above categories divided by the total number of students who belong

⁵⁶There might be a valid concern about the sufficient amount of variation in the average achievement of incoming students to estimate peer effects. The variance decomposition of the average test score of new students in class shows that about 60% of the variation in the average quality of new students is explained by between school variation. This variation is absorbed by the school fixed effects in the regressions. From 14% to 22% of the variation in the proportions of ESL students, foreign-born students or students who did not learn English at home is accounted for by the within school variation. For gender, approximately the same amount of variation comes from between and within school differences in the composition of peers. The large variation in the average quality of peers explained by the between school variation might be surprising, but this is not the same as variation in the individual achievement level. As expected, at individual level, about 80% of variation is explained by within school differences in students' test scores.

⁵⁷I analyzed the impact of the share of the new classmates on incumbent's achievement. Results are presented in Table 6A in Appendix. Overall, a higher turnover in a class has a negative impact on contemporaneous achievement with girls affected more than boys. The average effect, however, masks the fact that students who are enrolled into English as a Second Language program and who did not learn English at home benefit from the high turnover of classmates.

to that category, by classroom. As seen from the graph, the distribution of the averages for incumbent students is approximately normal, which suggests that grouping by abilities is not the apparent feature of the data.⁵⁸ If new students were matched to classmates in order to preserve the average achievement in each class, then the two distributions would be approximately the same. According to Figure 1, this is not the case, since the distribution of averages for new students does not exhibit any particular pattern except it peaks at the median. For comparison, Figure 2 presents the distribution of girls' shares for new and incumbent students. These distributions are similar to each other, as expected. In the online survey, school principals confirmed that they strive to have gender balance in classrooms.

To learn about the assignment of new students to classes, I estimate a series of regressions to predict Grade 3 test score of incoming student based on the average lagged achievement level in the classroom. Using the results of regressions reported in Table 2, I can test a number of important hypotheses about the placement of new students within a school.⁵⁹ If placement of new students into classes within a school is indeed random and not based on their abilities or other unobserved characteristics, then I should not observe correlation between the lagged achievement of a new student and the average achievement in the classroom. Column (1) of Table 2 demonstrates that Grade 3 math test score of a new student is not related to the average score of his/her new classmates. Nor is it related to other observed characteristics of other students in that classroom: share of girls, ESL and foreign born students. I also include interactions of the average achievement with indicators of a new student's background - gender, foreign born, English as a second language and whether

⁵⁸In order to learn whether the ability-tracking is a common practice in Ontario elementary schools, I also regressed the standard deviation of a class test scores on a standard deviation of school test scores accounting for school and cohort fixed effects and class size. The coefficient estimates is 0.9 which implies that the distribution of test score within class is not much tighter compared to the school distribution. In the presence of ability tracking, we would expect to observe more homogeneous distribution of achievement at a class level than at a school level.

⁵⁹Ammermueller and Pischke (2009) conduct a series of Pearson chi2 tests to show that the assignment of fourth grade students in European schools is plausibly random. They test whether the characteristics of classmates are independent conditional on school level value of that characteristics and perform a Pearson chi2 test for each school with more than one class. I turn to a different strategy because I use a different measure of peer quality and can directly identify new students in a class or in a school.

student learned English at home. Only one coefficient is significant; it implies positive correlation between the lagged achievement of a new student and a share of students who did not learn English at home.

In column (2) I test whether the same result holds for incumbent students in a classroom. To do that, I randomly draw a sample of incumbent students equal to the size of the cohort of new students in Grade 6, and designate them as new. I then estimate the same regression as in column (1). The correlation between the lagged test score of a student and the average score of classmates is negative. This implies that incumbent students are in fact balanced in classes based on their observed achievement and confirms that the sorting of students within school is non-random.⁶⁰

Next, I estimate the relationship between the probability to be placed in a given class and the average achievement of incumbent students. Columns (3) and (4) report results for two different definitions of a new student. In Column (3) a new student is defined as someone who entered in Grade 6. In Column (4), new students are all students who entered between Grades 4 to 6. The concern here is that new students may be allocated to classes not based on their own ability, but on the average achievement of students in that class. For instance, all new students are assigned to a class with the highest average achievement to help incoming students better adapt to a new school. The results in columns (3) and (4) suggest that new students do not seem to be tracked to specific classes.⁶¹ There is no correlation between having a new student in a class and the average lagged achievement of incumbent students in that class.

The last four columns of Table 2 report the coefficients from regressions that relate background characteristics of a new student to the average characteristics in a classroom. In general, results in Table 2 support the findings from the field about how new students are allocated to classes.⁶² School heads indicated that they try to balance gender mix in

⁶⁰I repeated the procedure 1000 times to ensure that the result is not generated by chance. Out of 1000 simulations, only in 7 cases the coefficient of interest was insignificant.

⁶¹The explanatory power of these regressions is also very low - the R^2 is at most 3%.

⁶²In the Appendix, I provide open-ended answers of school principals to my questions about the place-

classes if possible, and I find that a new girl is more likely to be assigned to a class with smaller share of girls (column 4). School principals also noted that they try to “bundle” English as a second language students in the same classroom to use teacher’s resources more efficiently. That is why it is not surprising to find in the data that classes with originally high fraction of ESL students are more likely to get a new classmate who also learns English as a second language (column 5). According to the results of the estimation, foreign born students are not matched to each other as school administration strives to create diversified classrooms and inclusive environment for all students.

The analysis of students’ assignment to classes combined with the information from the school heads implies that the allocation of new students into classes within a school can be considered plausibly random based on the observed student’s characteristics. Thus, the identification strategy based on the variation in the quality of incoming students in class should yield the reliable estimates of peer effects in elementary schools. The next section describes the results of the empirical analysis.

4 Results

4.1 Linear-in-means model

In this section, I present the results of the estimation of equation (4). The main results are described using mathematics test score; results with reading and writing tests are also discussed below and presented in Appendix Tables 2A and 3A respectively. The coefficients estimates are different across three subjects, while the significance and structure of the effect are the same.

ment of new students, and they confirm that administration of schools as a rule does not know the past achievement of students and allocate incoming students based on the class size consideration, special needs of the students - such as English as a second language, gifted. Schools also have rules to allocate students into classes for the next school year. While these rules are the prerogative of the school administration, the common guidelines include (1) balance of low, average and high achievers; (2) gender balance, balance of students with special needs and behaviour. Extract from the rules in one of the representative elementary school is provided in Appendix.

As has been discussed in section 2, the within school OLS estimates of peer effects are likely to be biased even after controlling for non-random selection into schools because of the correlation between the unobserved determinants of individual achievement and the average achievement level of classmates. To understand the direction of the bias, I turn to Table 3 which presents the OLS estimates from the regression of individual test score in Grade 6 on controls and peer variable which is defined in different ways. Column (1) in Table 3 reports the within school OLS estimate from the regression of individual test score in Grade 6 on the average contemporaneous achievement level of *all* students in class. The large and statistically significant coefficient implies that the individual test score and the average test score of classmates are positively related and documents that similar students tend to attend the same school. This specification does not include school and cohort fixed effects, so it is reasonable to assume that the OLS estimate of peer effect is biased upward because of the positive selection of parents into neighborhoods and students into schools.

In the specification in column (2), school fixed effects account for non-random selection of students into schools, and year and school-by-year fixed effects absorb cohort specific and time variant trends in achievement levels across schools. The peer effects coefficient in column (2) is significantly smaller than its counterpart in column (1) and precisely estimated. The difference in the magnitude of the coefficients in the first two columns confirms that selection into schools is positive and naive ordinary least squares estimation overstates the impact of classmates. The coefficient estimate from the fixed-effects specification is still likely to be biased. The identification with school fixed effects comes from the within school variation in peers' average achievement but the assignment of students into classrooms within school is non-random. To account for non-random placement of students into classes, specification in column (3) estimates a reduced-form relationship between an individual test score in Grade 6 and the average lagged test score of new classmates. The reduced-form coefficient can be thought of as direct impact of the average ability of *new* classmates on the outcome of the incumbent students.

While the magnitude and significance of the coefficients in columns (2) and (3) is the same, the interpretation is different. The coefficient in column (2) reports a positive correlation between a student’s and classmates contemporaneous achievement. The estimate in column (3) represents the causal impact of new peers on individual outcome, conditional on assumption of plausibly random initial assignment of new students into classes. For comparison, regression in column (4) uses the average lagged test score of only “old” classmates, i.e. those students who stayed in the same school over the observed period. The estimate is different from the estimate in column (2) and points out to the potential endogeneity problem when using the lagged test scores of incumbent peers to estimate ability spillovers. The last column in Table 3 shows the OLS estimate when the peer variable is defined as the average lagged test score of all students in class - new and old peers. The estimate is large and similar in magnitude to the coefficient when peer group is defined as incumbent students only. The reason why the estimates in two last columns are similar is because the distribution of the overall lagged class average does not change significantly by adding scores of new students since the fraction of new students is small.⁶³ The significant difference between the estimates using contemporaneous and lagged test score of the peer group shows that the current peer group is endogenous to individual achievement.

The swings in the coefficient estimates in Table 3 implicitly support the idea of using new peers as an instrument for the average achievement of all classmates. These results also confirm the endogeneity concern with the peer variable defined as the average test score of all classmates independent of whether it is a lagged or contemporaneous outcome. Given these preliminary results, I move on to estimate the effect of the average achievement of all classmates on individual test score using the IV strategy.

Table 4 presents the results of IV estimation as described in section 2. The first two columns repeat the OLS and reduced-form estimates from Table 3 for comparison with IV estimates. As noted before, the OLS estimate is likely to be biased. The instrumental

⁶³Figure 2 in Appendix illustrates this point by plotting the distribution of the averages for two definitions of the peer group.

variables strategy should provide the insights into the direction and magnitude of the bias if the identification assumption is valid. Column (3) presents the coefficients from the first stage of the 2SLS procedure (equation 3) and implies that there is a strong first-stage relationship between the average peer achievement and average lagged test score of new classmates. In Table 7A in Appendix I present additional evidence that new students to a class have an impact on overall class achievement. Class-level regressions in Table 7A demonstrate that new students in class significantly shift the average achievement of other students in that classroom in Grade 6. Controlling for the average lagged test score and variability of test scores of incumbent students, class size and share of new students, I find that the contemporaneous achievement of all students in class is positively correlated with the quality of new students.

The IV estimate in column (4) is more than six times larger than its OLS counterpart and statistically significant. The larger IV estimates compared to the OLS estimates imply that the OLS coefficient is biased downward. If students were grouped in classes by abilities, i.e. were matched with students with the similar achievement level, then the OLS would yield a positive effect even if there was none, and the magnitude of the coefficient is expected to go down when instrumental variable is introduced. If students were initially mixed in heterogeneous classrooms, then the correlation between own ability and ability of the classmates would be negative. The IV estimate under the second scenario would be larger than the OLS estimate. This is exactly what is observed in Table 4 when we move from column (1) with OLS estimates to column (4) which presents IV estimates. If new students were allocated to classrooms in a systematic way - mixed or matched by abilities, then I should not have observed a large swing in the magnitude of the coefficient when moving from OLS to IV estimates. In other words, the instrument would not add new information to the identification of peer effect compared to the OLS estimate. The significant difference between the OLS and IV estimates indicates that the instrument is working. The result of the Hausman test confirms that contemporaneous average achievement is endogenous and

that IV estimates systematically differ from their OLS counterparts with p-value of the test of 0.002. The magnitude of the IV estimate shows that for every standard deviation increase in the average test score of classmates, an individual math test score increases by a quarter of a standard deviation.⁶⁴

The magnitude of the effect found in this paper is relatively large compared to other studies of peer effects in education. One reason why the estimate in my study is large is the choice of the relevant peer group. As a rule, studies that examine peer effects at a school or grade level find modest peer effects. A few papers which were able to estimate both classroom and school level effects, find that the size of the effect becomes larger when moving from grade or school to class level.⁶⁵ One potential explanation for why it might be the case is that the peer quality matters more *within* the classroom because the distribution of students' abilities influence teaching style and teacher's effort and because students in elementary school spend more time with immediate classmates as opposed to schoolmates. Conversely, if school peers matter more then it must be that the interactions happen at a more aggregate level through the endogenous sorting into peer group. I further investigate the difference between classroom and grade level peers in section 4.3.

The finding that the average test score of classmates has a positive and statistically significant impact on individual test score is interesting, but does not inform about the efficient allocation of students in classes because assumes a zero-sum game from switching peers across classrooms. The logical step would be to test the linear-in-means model of peer effects.

⁶⁴The magnitude of the effect is computed as an effect on individual Grade 6 test score due to a one standard deviation change in the average ability of current classmates. The magnitude of the effect is comparable to the largest estimates reported in the literature. The majority of the studies of peer effects in education use the variation in the background characteristics of school- or classmates to estimate the magnitude of the interactions, and not the direct measure of ability - test score. Among those who use test score data, the effects of a 1 sd. deviation increase in peer score range from zero (Lavy, Silva and Weinhardt (2012)) to 0.33 of st.deviation in Imberman, Kugler and Sacerdote (2012).

⁶⁵See, for instance Vygdor and Nechyba (2007) for North Carolina schools and Burke and Sass (2011) for students in Florida public schools who find larger effect at a classroom level compared to school and/or grade level. There are exceptions to this pattern - Betts and Zau (2003) find larger effects at grade and school level as opposed to classroom.

Equation (5) describes the conventional linear-in-means model and introduces heterogeneity by initial achievement:

$$Y_{icst} = \beta_0 + \sum_{j=1}^4 \beta_{1j} \bar{Y}_{(-i)cs,t} \times D_{ics,t-1}^j + \varepsilon_{icst} \quad (5)$$

where the dummy variable D_{icst}^j indicates the lagged achievement level of student i . The endogenous terms $\bar{Y}_{(-i)cs,t-1} \times D_{ics,t}^j$ 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.

If the average impact of classmates is the same for everyone independent of abilities approximated here by the lagged achievement level, then the null hypothesis of the equality of β_{1j} should not be rejected. According to the results in columns (5)-(8) of Table 4, the data do not provide support in favour of linear-in-means model. The null hypothesis of the equal effect of the average quality of classmates on students with different initial level of achievement is rejected at 1% significance level. The impact of classmates is increasing in own ability which implies that high achievers benefit the most from the increase in the average quality of their peers. This finding is consistent with the monotonicity model of peer effects.⁶⁶ The monotonicity model assumes that the impact of peers is increasing in own achievement. The result also suggests that the conventional linear model might not be the best to explain the social interactions in elementary school.

Another dimension of the heterogeneity in response to peer interactions is the gender of a student. Previous studies documented the differences in the effect on girls and boys from the gender class composition (Hoxby, 2000; Lavy and Schlosser, 2011), and from the shares of bad and good students (Lavy, Silva and Weinhardt, 2012). Table 5 reports the results of the estimation for two sub-samples of incumbent students - boys and girls. Overall, I find

⁶⁶Various models of social interactions in school are described in Hoxby and Weingarth (2005) and Sacerdote (2011).

positive and statistically significant effect of the average peer ability on both boys and girls, but the impact on girls is larger. Translating the difference into standardized measures, a one standard deviation increase in peers' test score implies 0.28 standard deviations increase in test score for girls, and 0.22 standard deviation increase for boys.⁶⁷ The impact is the largest for incumbent girls who are at the top of ability distribution themselves. It is not surprising to find different impact on boys and girls since a growing number of studies shows that girls are more affected than boys by interventions and education inputs.⁶⁸ Also, peer effects might work through different channels and in different ways for boys and girls.

4.2 School peers vs classroom peers

In this section, I contrast the class level estimates of peer effects described above with the school level estimates obtained using the same strategy. I do that in order to understand whether the definition of the relevant peer group may be the reason for small or insignificant estimates of peer effects in elementary education in previous studies.⁶⁹

One of the main advantages of the Ontario test scores data is that it allows to identify peers not only at a school or grade level but also at a classroom level. The definition of the relevant peer group plays important role in the identification of ability spillovers. As a rule, the peer group in previous studies was identified at a grade or cohort level. The common rationale for not using classroom as a peer group builds around the argument that students are systematically assigned to classes in an unobserved way. While this is true in general, and school administration as well as parents are likely to influence the placement of students to classes, the allocation of *new* students is less likely to be systematic based

⁶⁷This result is consistent with Lavy, Silva and Weinhardt (2012) who find that girls benefit from the increase in top 5% of classmates, while boys do not. However, they do not find evidence of the average effect on either gender. Lavy and Schlosser (2011) report slightly smaller effect on girls than on boys from the increase in share of female classmates.

⁶⁸Angrist and Lavy (2009) find no effect of a cash incentive program for low achievers on boys and large effect on girls. Angrist, Lang and Oreopoulos (2009) also find no effect of an incentive based program on male college freshman, but large effect on female.

⁶⁹Halliday and Kwak (2012) address the question about the relevant peer group using Add Health data and show that incorrectly defining the peer group may lead to underestimates of peer effects in educational context. The authors contrast the estimates for grade level peers with self-nominated friend's group.

on their achievement. There are at least two reasons for assuming that. First is that the school administration has an imperfect information about the abilities of a new student. The EQAO test scores of new students are not recorded in the report card, and school principals mentioned in the survey and interviews that as a rule they do not know the abilities of the child before s/he enters the school. Second, parents might not have enough information about classrooms and teachers to make a request to assign their child to a specific teacher. As has been discussed in section 3, the empirical evidence in this study support the assumption of plausibly random assignment of *new* students into classrooms. There are no direct evidence that the placement of new students is related to the average lagged ability in the classroom conditional on observed background characteristics and lagged test score of a new students. In this paper, I can go beyond the school and grade level peer effects and estimate ability spillovers at the level of the classroom because the within school sorting is handled using new peers as an instrument. Also, I can compare the estimates for the broad definition of the peer group - school, with the estimated impact of peers at a classroom level. I expect such comparison to shed the light into the question about the relevant peer group for elementary students.

I construct an instrument at a school-cohort level as the average lagged test score of all new students in a school and estimate equation (4) with school level peers instead of classroom level peers. Results of the estimation are presented in Panel B of Table 6 and contrasted with the similar estimates from Table 4. The OLS estimate for the school-level peers are larger than the estimate for the classmates. As discussed before, the OLS estimate in class level regression appear to be biased downward and the direction of the bias suggest that school administrators are likely to balance students by abilities. In class level regressions I compare classes with different average achievement of incoming students. In school-level regressions, I compare the average ability of incoming students across adjacent cohorts of six grades. As before, school fixed effects control for any unobserved differences between schools, school-by-year fixed effects absorb time-variant differences and year fixed

effects account for time trends. The reduced form coefficient in column (2) is small and insignificant. The first stage relationship between lagged achievement of new students and contemporaneous achievement of all students in a grade is highly significant implying that the lagged achievement of incoming students predicts the contemporaneous test score of all schoolmates. The coefficient in column (4) suggests that there is no ability spillovers from school level peers and that the OLS estimate overstates the impact of schoolmates. It is not surprising to find a positive bias in school-level estimates.

Combined together, estimates in columns (2)-(4) suggest that the instrument is working but the endogeneity in school level estimates has a different nature compared to classroom level peers. The story in Table 6 unfolds as follows. Residential choice of parents determines the sorting of children into schools. According to OLS estimate this sorting is positive and similar children attend the same school which is confirmed by both school and classroom level estimates. Then, school administration decides about the optimal classroom assignment. The IV estimates imply that students are mixed in classrooms to create balance of abilities and other characteristics. For grade level peers, school is the last level of sorting as students can not be randomly allocated to grade and this is primarily based on their age. At a grade or school level new peers as an instrument solve the reflection problem. Instrument takes care of the simultaneity of the outcomes between schoolmates and it appears to be no spillover at a school level. At a classroom level, using new peers as an instrument also overcomes non-random selection into classes.

One reason for why school level estimates are smaller and non-significant as opposed to class level estimates is the definition of the relevant peer group for elementary school students. Children spend at least six hours a day for a minimum of 194 days a year with their classmates. The time when they are exposed to other students in school, for instance during the recess period, is significantly shorter. The majority of the interactions, and especially interactions that involve learning, happen within the classroom. In the same

way as individuals tend to make friends with those who live in their close proximity⁷⁰, it is reasonable to assume that in elementary school students are more likely to become friends with their immediate peers - classmates. This might not be true in higher education, where students are more mobile both within a school and their neighborhood. The relevant peer group for older students is likely to expand beyond the classroom. Older students also more likely to make friends based on the similarity in background characteristics, rather than on belonging to the same classroom. There are no empirical evidence about the pattern of friendship formation for elementary school students. The findings in this section, however, allow to assume that for younger children the relevant peer group is most likely to be their immediate classmates rather than all students in their school or grade.

These results are consistent with a small number of papers that estimated peer effects at a classroom level. Ammermueller and Pischke (2009) provide evidence of peer effects for fourth graders in six European countries exploiting plausibly random assignment of students into classrooms within schools. They find modestly large effects of peers on academic achievement after correcting for the measurement error in the peer variable. Cooley (2010a) uses a change in the accountability rules in North Carolina that presumably affected the level of effort exerted by students at different ends of ability distribution to estimate peer effects in the classroom and finds large and significant effects. Finally, Burke and Sass (2011) exploit longitudinal data from Florida to recover significant spillovers from unobserved peer ability among elementary students.

4.3 Robustness Checks

In this section, I present a number of empirical tests to show the robustness of the identification strategy to potential threats. I start with replicating the results in Table 4 with an instrument defined over all new students who entered a school between the start of Grade 4 and Grade 6. Since I cannot track transition between classrooms over the three year, all

⁷⁰See, for instance Mouw and Entwisle (2009) for the evidence that distance affects friendship.

I know about all new students is that they are new to a school and took Grade 3 test in another school. Thus, the new students who are new schoolmates (i.e. those who entered in Grade 4 and Grade 5) may have been classmates with incumbent students before. Moreover, these new students may have been placed to classroom based on the observed abilities in their new school. The non-systematic sorting shown to exist in elementary school in section 3 would lead to endogeneity of the peer variable defined over all new students and bias in the peer effect estimate. According to Table 2A in Appendix, the impact of new students remains to be positive and significant. High-achievers among incumbent students are the ones who benefit the most from the increase in the average ability classmates. The effect of classmates achievement varies with own ability as confirmed by the test of linear-in-means model. Overall, the findings using all new students in class are consistent with the main results in the study.

Tables 3A and 4A replicate the results from Table 4 using reading and writing test scores instead of mathematics test score. These results demonstrate that the definition of the outcome measure is unlikely to drive the positive and significant peer effect in Table 4. The estimates of the impact of peers on both reading and writing test score are smaller in magnitude than the effect on math test score, but positive and significant. This finding is consistent with other studies that also looked at peer effects for different subjects. For instance, in Imberman, Kugler and Sacerdote (2012) the effect on math test score from the hurricane evacuees is larger than the impact for reading test scores for both elementary and middle school students. Burke and Sass (2011) find positive and significant effect of classmates on math test score gain and no effect on average on the reading test score gain.

Next, I restrict sample to students who are enrolled in schools with only one Grade 6 class. For schools with only one class, the selection into school is the same as selection into class. The identification of peer effects in this case comes from the variation in the average ability of incoming students across cohorts within school after the non-random selection into schools is absorbed by the school-fixed effects. The remaining concern is that the test

scores of classmates are determined simultaneously such that the unobserved determinants of individual achievement are correlated with the average test score of classmates. To overcome this simultaneity, I use lagged test score of new students which were determined before they switched the school. Table 5A in Appendix replicates Table 4 using the restricted sample of schools with one Grade 6 class. The results show similar pattern - the IV estimate is larger than the OLS estimate and statistically significant. The OLS and IV estimates are systematically different from each other as confirmed by the Hausman test (the p-value of the test is 0.002).

I have also experimented with different samples as described in Appendix Table 1A to check whether the main results are driven by the new students systematically assigned to classrooms. I did not find any significant difference in the estimates when I exclude ESL students, or students enrolled into a French immersion program, or foreign born students.

5 Conclusion

In this paper, I combine a rich data set from Canadian province of Ontario with institutional features of public education system to estimate peer effects in elementary school and provides evidence in favor of sizable ability spillovers among students in the classroom.

Using the unique features of the data and instrumental variables strategy I am able to overcome selection and simultaneity problems that hinder the estimation of peer effects. I demonstrate that the assignment of new students to classes within a school is plausibly random and use the average lagged achievement of new classmates to recover the impact of peers on individual achievement.

I find positive and significant effects of good peers that operate at a classroom level. The marginal effect of a one standard deviation in peers quality leads to a 0.25 standard deviation increase in individual achievement. These estimates are robust to the definition of the peer group, outcome measure and sample selection. I find that peer effects are heterogeneous in

nature and monotone in student's own ability. The impact of peers depends not only on the average achievement in a peer group but varies with individual ability. While all students in class benefit from good peers, the impact is largest for high-achievers. Another dimension of heterogeneity is the effect of peers on girls and boys. I find larger average effect on girls compared to boys. Within girls and boys samples, the structure of the impact is the same as for the entire sample with high-achieving students benefiting more from good classmates. I also compared the impact of immediate classmates with the impact of school- or grade level peers. I found no effect of schoolmates on individual achievement. I argue that the relevant peer group for elementary students is a classroom as opposed to all students in the same grade or school and that aggregation of the peer group to a school or grade level masks the peer effects even if they are present.

While I argue that the identification strategy in this paper together with a unique data set mitigates some of the problems with estimation of peer effects, I am aware of the limitations of this study. The main limitation is in the observational nature of the data which makes informing policy based on these estimates a risky task. This study does, however, inform us about the nature of selection in schools and the way school administrators form the classes. It also provides credible estimates of ability spillovers given the way classrooms are organized.

References

- [1] Ammermueller, A., Pischke, J.-S. (2009). Peer Effects in European Primary Schools: Evidence from the Progress in International Reading Literacy Study. *Journal of Labor Economics*, 27(3), 315–348.
- [2] Angrist, J. and Lang, K. (2004). Does school integration generate peer effects? Evidence from Boston's Metco program. *American Economic Review*, 94, 1613–1634.

- [3] Angrist, J. and Lavy, V. (2009). The effects of high stakes high school achievement awards: evidence from a randomized trial. *American Economic Review*, 99(4), 1384-1414.
- [4] Angrist, J., Lang, D, and Oreopoulos, P. (2009). Incentives and services for college achievement: evidence from a randomized trial. *American Economic Journal: Applied Economics*, 1(1), 1-28.
- [5] Arcidiacono, P. and Nicholson, S. (2005). Peer effects in medical school. *Journal of Public Economics*, 89, 327–350.
- [6] Baker, M. (2012). Industrial Actions in Schools: Strikes and Student Achievement. *Canadian Journal of Economics*, forthcoming.
- [7] 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.
- [8] Betts, J. and Shkolnik, J. (2000a). The Effects of Ability Grouping on Student Achievement and Resource Allocation in Secondary Schools. *Economics of Education Review*, 19, 1-15.
- [9] Betts, J. and Shkolnik, J. (2000b). Key difficulties in identifying the effects of ability grouping on student achievement. *Economics of Education Review*, 19, no.1, 21-26.
- [10] Betts, J. and Zau, A. (2004). Peer groups and academic achievement: Panel evidence from administrative data. Unpublished manuscript.
- [11] Betts, J. (2011). The Economics of tracking in education. Handbook of Economics of Education, Volume 3, Hanushek, E., S. Machin and L. Woessmann, eds., Amsterdam: North Holland, 341-381.
- [12] Black, S., Devereux, P. and Salvanes, K. (2013). Under Pressure? The effects of peers on outcomes of young adults. *Journal of Labor Economics*, 31(1), 119-153.

- [13] Blume, L., Brock, W., Durlauf, S. and Ionnides, Y. (2010). Identification of social interactions. Handbook of Social Economics, J. Benhabib, A.Bisin, and M.Jackson, eds., Amsterdam: North Holland.
- [14] Bobonis, G. and Finan, F. (2009). Neighborhood peer effects in secondary school enrollment decisions. *Review of Economics and Statistics*, 91(4), 695-716.
- [15] Boozer, M.A. and Cacciola, S.E. (2001). Inside the black box of project star: Estimation of peer effects. *Economic Growth Center Discussion Paper #832*.
- [16] Boucher, P., Bramouille, Y, Djebbari, H. and Fortin, B. (2012). Do peers affect student achievement ? Evidence from Canada using group size variation. *Journal of Applied Econometrics*, forthcoming.
- [17] Burke, M.A. and Sass, T.R. (2013). Classroom peer effects and student achievement. *Journal of Labor Economics*, 31(1), 51-82.
- [18] Card, D., Dooley, M.& Payne, A. (2010). School Competition and Efficiency with Publicly-Funded Catholic Schools. *American Economic Journal: Applied Economics*, 2(4).
- [19] Carrell, S., Fullerton, R.L. and West, J.E. (2008). Does your cohort matter? Measuring peer effects in college achievement. *Journal of Labor Economics*, 27(3), 439-464.
- [20] 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.
- [21] Carrell, S., Sacerdote, B. and West, J. (2012). From Natural Variation to Optimal Policy? An Unsuccessful Experiment in Using Peer Effects Estimates to Improve Student Outcomes. *NBER Working paper # 16865*.

- [22] Cooley, J. (2010a). Desegregation and the Achievement Gap: Do Diverse Peers Help?. Working paper, Department of Economics, University of Wisconsin-Madison.
- [23] Cooley, J. (2010b). Can Achievement Peer Effect Estimates Inform Policy? A View from Inside the Black Box. Working paper, Department of Economics, University of Wisconsin-Madison.
- [24] Deaton, A. (2010). Instruments, randomization, and learning about development. *Journal of Economic Literature*, 48, 424-455.
- [25] 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.
- [26] 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.
- [27] Epple, D., and Romano, R. (1998). Competition between private and public schools, vouchers, and peer-group effects. *American Economic Review*, 88, 33-62.
- [28] Evans, W.N., Oates, W.E. and Schwab, R.M. (1992). Measuring peer effects: A study of teenage behavior. *Journal of Political Economy*, 100, 966-991.
- [29] Friesen, J. and Krauth, B. (2007). Sorting and inequality in Canadian schools. *Journal of Public Economics*, 91(11-12), 2185-2215.
- [30] Friesen, J. and Krauth, B. (2010). Sorting, peers and academic achievement of aboriginal students in British Columbia. *Canadian Journal of Economics*, 43(4).
- [31] Friesen, J. and Krauth, B. (2011). Ethnic enclaves in the classroom. *Labour Economics*, 18(5), 656-663.

- [32] 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.
- [33] Gibbons, S. and Telhaj, S. (2012) Peers and achievement in England's secondary schools. *SERC Discussion Paper #1*.
- [34] Gould, E., Lavy, V. and Passerman, D. (2009). Does immigration affect the long-term educational outcomes of natives? Quasi-experimental evidence. *Economic Journal*, 119, 1243–1269.
- [35] Halliday, T.J and Kwak, S. (2012). What is a peer? The role of network definitions in estimation of endogenous peer effects. *Applied Economics*, 44(3), 289-302.
- [36] Hanushek, E.A., Kain, J.F., Markman, J.M. and Rivkin, S.G. (2003). Does peer ability affect student achievement? *Journal of Applied Economics*, 18, 527-544.
- [37] Hanushek, E.A., Kain, J.F. and Rivkin, S.G. (2009). New evidence about Brown v. Board of Education: The complex effect of school racial composition on achievement. *Journal of Labor Economics*, 27(3), 349–383.
- [38] Hoffer, T. (1992). Middle school ability grouping and student achievement in science and mathematics. *Educational Evaluation and Policy Analysis*, 14, no.3, 205–227.
- [39] Hoxby, C. (2000). Peer effects in the classroom: Learning from gender and race variation. *NBER Working Paper 7867*.
- [40] Hoxby, C., Weingarth G. (2006). Taking race out of the equation: School reassignment and the structure of peer effects. Harvard University, unpublished manuscript.
- [41] 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.

- [42] Kang, C. (2007). Classroom peer effects and academic achievement: quasi-randomization evidence from South Korea. *Journal of Urban Economics*, 61, 458-495.
- [43] Lavy, V., Schlosser, A. (2011). Mechanisms and impacts of gender peer effects at school. *American Economic Journal: Applied Economics*, April 2011.
- [44] 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.
- [45] 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.
- [46] Liu, X., Patacchini, E. & Zenou, Y. (2011). Peer Effects in Education, Sport, and Screen Activities: Local Aggregate or Local Average?, *CEPR Discussion Paper No. DP8477*
- [47] Manski, C. (1993). Identification of endogenous social effects: the reflection problem. *Review of Economic Studies*, 60(3), 531–542.
- [48] Micklewright, J., Schnepf, S.V. & Silva, P. (2012). Peer effects and measurement error: The impact of sampling variation in school survey data (evidence from PISA). *Economics of Education review*, 31(6), 1136–1142.
- [49] Moffitt, R. (2001). Policy interventions, low-level equilibria and social interactions. In: Steven N. Durlauf and Peyton H. Young (*Eds*), *Social Dynamics*, Chapter 3. The MIT Press.
- [50] Ted Mouw, T. and Barbara Entwisle, B. (2006). Residential segregation and interracial friendship in schools. *American Journal of Sociology*, 112(2), 394-441
- [51] Rees, D, Brewer, D. and Argys, L. (2000). How should we measure the effect of ability tracking on student performance? *Economics of Education Review*, 19, no.1, 17–20.

- [52] 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.
- [53] Sacerdote, B. (2001). Peer effects with random assignment: Results for Dartmouth roommates. *Quarterly Journal of Economics*, 116, 681–704.
- [54] 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.
- [55] Slavin, R. (1987). Ability grouping and student achievement in elementary schools: a best-evidence synthesis. *Review of Educational Research*, 57, no.3, 293–336.
- [56] Slavin, R. (1990). Achievement effects of ability grouping in secondary schools: best-evidence synthesis. *Review of Educational Research*, 60, no.3, 471–499.
- [57] 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.
- [58] Whitmore, D. (2005). Resource and peer impact on girls' academic achievement: Evidence from a randomized experiment. *American Economic Review Papers and Proceedings*. 95(2), 199–203.
- [59] Zimmerman, D.J. (2003). Peer effects in academic outcomes: Evidence from a natural experiment. *Review of Economics and Statistics*, 85, 9–23.

Tables and Figures

Figure 1:

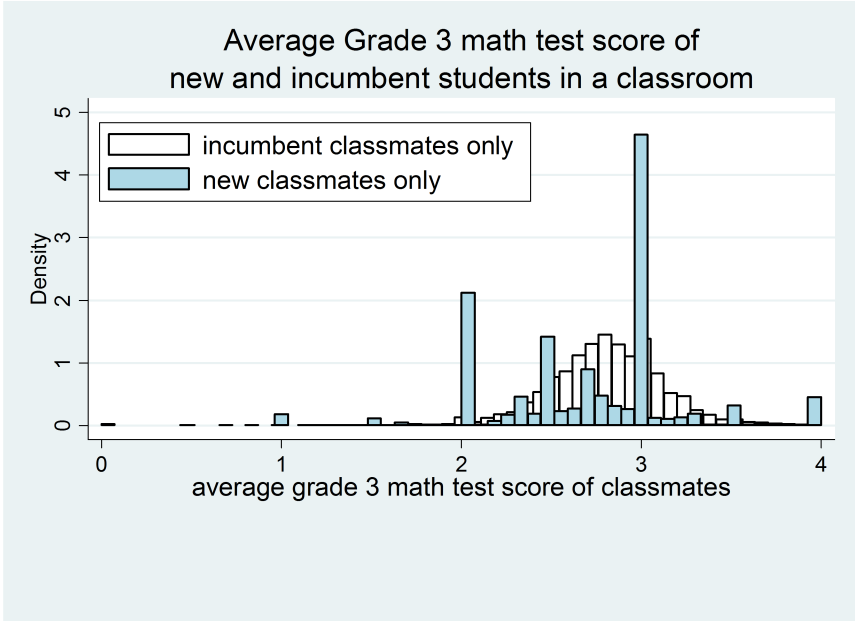


Figure 2:

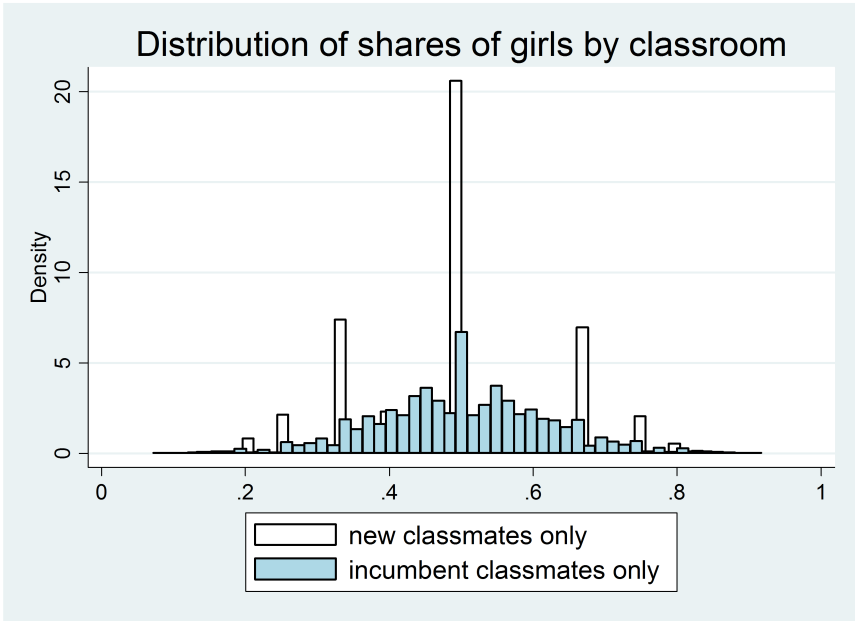


Table 1. Summary Statistics

	2008		2009		2010	
	incumbent	new	incumbent	new	incumbent	new
Math test score, Grade 3	2.76 (0.67)	2.64 (0.72)	2.81 (0.67)	2.69 (0.71)	2.81 (0.67)	2.68 (0.71)
Math test score, Grade 6	2.74 (0.74)	2.56 (0.78)	2.78 (0.76)	2.59 (0.81)	2.75 (0.78)	2.56 (0.81)
Level 1 in mathematics G6, %	0.03 (0.17)	0.05 (0.21)	0.03 (0.16)	0.04 (0.20)	0.03 (0.16)	0.04 (0.21)
Level 2 in mathematics G6, %	0.27 (0.45)	0.33 (0.47)	0.24 (0.43)	0.31 (0.46)	0.25 (0.43)	0.32 (0.47)
Level 3 in mathematics G6, %	0.60 (0.49)	0.54 (0.50)	0.61 (0.49)	0.56 (0.50)	0.61 (0.49)	0.55 (0.50)
Level 4 in mathematics G6, %	0.10 (0.30)	0.08 (0.27)	0.11 (0.32)	0.09 (0.29)	0.12 (0.32)	0.09 (0.28)
Average G3 math test score of classmates	2.75 (0.30)	2.72 (0.33)	2.80 (0.30)	2.77 (0.33)	2.80 (0.30)	2.77 (0.33)
Average G6 math test score of classmates	2.73 (0.40)	2.66 (0.43)	2.76 (0.38)	2.70 (0.43)	2.73 (0.41)	2.67 (0.44)
Reading score, Grade 3	2.61 (0.76)	2.46 (0.82)	2.66 (0.74)	2.53 (0.79)	2.63 (0.75)	2.49 (0.82)
Reading score, Grade 6	2.75 (0.67)	2.62 (0.73)	2.83 (0.67)	2.69 (0.72)	2.85 (0.66)	2.73 (0.72)
Writing score, Grade 3	2.70 (0.63)	2.58 (0.66)	2.74 (0.59)	2.54 (0.63)	2.74 (0.59)	2.64 (0.62)
Writing score, Grade 6	2.81 (0.61)	2.71 (0.65)	2.81 (0.59)	2.71 (0.61)	2.83 (0.59)	2.72 (0.60)
Female, %	0.50 (0.50)	0.49 (0.50)	0.49 (0.50)	0.49 (0.50)	0.49 (0.50)	0.49 (0.50)
English as a second language (ESL), %	0.02 (0.14)	0.03 (0.17)	0.02 (0.15)	0.04 (0.19)	0.03 (0.16)	0.04 (0.20)
Born outside Canada, %	0.07 (0.25)	0.12 (0.32)	0.08 (0.27)	0.12 (0.32)	0.08 (0.26)	0.12 (0.32)
Did not learn English at home, %	0.15 (0.36)	0.20 (0.40)	0.16 (0.36)	0.19 (0.39)	0.16 (0.36)	0.20 (0.40)
Number of observations	66,601	9,538	66,242	9,613	66,874	10,079

Note: Incumbent students are those who did not change school between Grade 3 and Grade 6, new students are students who moved to a new school after Grade 3 and before taking Grade 6 math test. Standard deviations in parentheses.

Table 2: Assignment of new students into classrooms

Dependent variable>>>>	Lagged test score of a new student (1)	Lagged test score of incumbent (2)	New student (=1) (3)	New student (=1) (4)	Gender of a new student (female=1) (5)	New student is in ESL program (6)	New student is foreign born (7)	New student did not learn English at home (8)
Average lagged math test score of incumbent students in class	0.054 (0.052)	-0.205*** (0.053)	0.000 (0.002)	-0.007 (0.006)	0.008 (0.031)	-0.000 (0.001)	0.026 (0.018)	-0.011 (0.019)
Fraction of female incumbent students in class	0.002 (0.067)				-0.088** (0.041)	0.604*** (0.106)		
Fraction of ESL incumbent students in class	-2.025 (3.782)							
Fraction of foreign-born incumbent students in class	0.132 (0.109)						0.004 (0.044)	
Fraction of students who did not learn English at home	0.247*** (0.099)							-0.001 (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)							
Foreign 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 probability 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 ability of classmates on individual test score - various definition of peers

Dependent variable>>>	Mathematics test score of incumbent student in Grade 6				
	OLS (1)	OLS (2)	OLS (3)	OLS (4)	OLS (5)
Average Grade 6 mathematics test score of all classmates	0.623*** (0.006)	0.059*** (0.013)			
Average Grade 3 test score of new classmates only			0.055*** (0.007)		
Average Grade 3 test score of "old" classmates				0.234*** (0.017)	
Average Grade 3 test score of all classmates					0.243*** (0.017)
Individual controls	Yes	Yes	Yes	Yes	Yes
School-year fixed effects	No	Yes	Yes	Yes	Yes
Year fixed effects	No	Yes	Yes	Yes	Yes
School fixed effects	No	Yes	Yes	Yes	Yes
Number of observations	199,717	199,717	199,717	199,717	199,717
R ²	0.37	0.33	0.28	0.28	0.28

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. Regressions cover years 2008-2010 and include individual controls (own test score in Grade 3, gender, English 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. Columns (1) and (2) show the OLS estimates of individual test score on the average Grade 6 test score of all students in class. In Column (3) the individual Grade 6 test score regressed on the average Grade 3 achievement of only new students in class. Column (4) shows the estimate from the regression of individual Grade 6 test score on the average Grade 3 test score of only incumbent students. In Column (5) the peer variable is defined is the average Grade 3 test score of all students in class.

Table 4. 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 achievement of incumbent student is:			
	OLS (1)	Reduced Form (2)	2SLS First Stage (3)	2SLS (4)	Level 1 (5)	Level 2 (6)	Level 3 (7)	Level 4 (8)
Average Grade 6 mathematics test score of classmates	0.059*** (0.013)		0.141*** (0.002)	0.419*** (0.053)	0.303*** (0.074)	0.316*** (0.059)	0.406*** (0.054)	0.455*** (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	Yes
Year fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
School fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
School-year fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Chi2 test of joint equality of coefficients								
Number of observations	199,717	199,717	199,717	199,717				
R ²	0.30	0.28	0.15	0.36				
						220.53***		
						199,717		
						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 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, English 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 5. Estimates of the impact of peers on boys and girls

Dependent variable>>>	Grade 6 mathematics test score of incumbent student				2SLS estimates of the effect of the average Grade 6 test score of classmates if initial achievement level of incumbent students is:			
	OLS (1)	Reduced form (2)	2SLS First stage (3)	2SLS (4)	Level 1 (5)	Level 2 (6)	Level 3 (7)	Level 4 (8)
Average Grade 6 math test score of classmates	0.093*** (0.019)	0.056*** (0.010)	0.088*** (0.002)	0.472*** (0.120)	0.307*** (0.107)	0.342*** (0.087)	0.446*** (0.081)	0.499*** *(0.084)
Number of observations	100,861	100,861	100,861	100,861		100,861		
Average Grade 6 math test score of classmates	0.063*** (0.020)	0.056*** (0.011)	0.098*** (0.002)	0.372*** (0.114)	0.347*** (0.108)	0.323*** (0.084)	0.387*** (0.074)	0.418*** (0.077)
Number of observations	98,856	98,856	98,856	98,856		98,856		
School, year and school-year fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

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. In Panel A - the sample comprises incumbent girls only, in Panel B - incumbent boys only. Regressions cover years 2008-2010 and include individual controls (own test score in Grade 3, gender, English as Second Language learner, Canadian born, and whether student learned English at home), average score of "old" peers; 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 6. Estimates of the impact of peers at school and classroom level

Dependent variable>>>	Grade 6 mathematics test score of incumbent student				2SLS estimates of the effect of the average Grade 6 test score of classmates if initial achievement level of incumbent students is:			
	OLS	Reduced form	2SLS First stage	2SLS	Level 1	Level 2	Level 3	Level 4
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
<i>Effect of classmates:</i> Average Grade 6 math test score of classmates	0.059*** (0.013)	0.055*** (0.007)	0.141*** (0.002)	0.419*** (0.053)	0.303*** (0.074)	0.316*** (0.059)	0.406*** (0.054)	0.455*** (0.055)
<i>Effect of schoolmates:</i> Average Grade 6 math test score of schoolmates	0.334*** (0.009)	0.005 (0.004)	0.055*** (0.001)	0.095 (0.071)	-0.106 (0.082)	-0.060 (0.079)	0.063 (0.078)	0.153* (0.080)
School, year and school-year fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Number of observations	199,717	199,717	199,717	199,717	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. 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: in Panel A - new students in the classroom, in Panel B - new students in grade/cohort. Regressions cover years 2008-2010 and include individual controls (own test score in Grade 3, gender, English as Second Language learner, Canadian born, and whether student learned English at home), average score of "old" peers; 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 Grade 6 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.