Do Peers Affect Student Achievement?
Evidence from Canada Using Group Size Variation

Vincent Boucher, Yann Bramoullé,
Habiba Djebbari, Bernard Fortin

Série Scientifique
Scientific Series

Montréal
Février 2010

CIRANO
Allier savoir et décision

Centre interuniversitaire de recherche en analyse des organisations

Do Peers Affect Student Achievement?
Evidence from Canada Using Group Size Variation[*]

Vincent Boucher [†], Yann Bramoullé [‡], Habiba Djebbari [§], Bernard Fortin [**]

Résumé / *Abstract*

Nous présentons une première application empirique d'une nouvelle approche développée par Lee (2007) pour estimer les effets de pairs dans un modèle linéaire-en-moyenne. Cette méthode permet de tenir compte des variables non-observées au niveau du groupe et de solutionner le problème de réflexion. Nous estimons les effets de pairs sur la performance scolaire (mesurée par les résultats aux épreuves du ministère de l'Éducation, du Loisir et du Sport) en Mathématiques, en Science, en Français et en Histoire dans les écoles secondaires du Québec. À cette fin, nous utilisons des méthodes de maximum de vraisemblance et de variables instrumentales. Nos résultats corroborent la présence d'effets de pairs. L'effet de pair endogène est positif, lorsqu'il est significatif. En particulier, une hausse d'un point dans la note moyenne de ses pairs accroît la note de l'élève de 0,5 en Français, de 0,65 en Histoire et 0,83 en Math (514). En outre, certains effets contextuels ont de l'importance. À partir de simulations Monte Carlo, nous trouvons qu'une grande variabilité dans la taille des groupes peut réduire les problèmes d'identification faible.

Mots clés : effets de pairs, performance scolaire, problème de réflexion.

We provide the first empirical application of a new approach proposed by Lee (2007) to estimate peer effects in a linear-in-means model. This approach allows to control for group-level unobservables and to solve the reflection problem. We investigate peer effects in student achievement in Mathematics, Science, French and History in Quebec secondary schools. We estimate the model using maximum likelihood and instrumental variables methods. We find evidence of peer effects. The endogenous peer effect is positive, when significant, and some contextual peer effects matter. Using calibrated Monte Carlo simulations, we find that high dispersion in group sizes helps with potential issues of weak identification.

Keywords: *peer effects, student achievement, reflection problem.*

Codes JEL : C31, I20, Z13

[†] Université de Montréal.
[‡] CIRPÉE and Department of Economics, Université Laval
[§] CIRPÉE, IZA and Department of Economics, Université Laval
[**] Corresponding author: CIRANO, CIRPÉE, and Department of Economics, Université Laval, Department of Economics, Université Laval, Quebec, Canada. Email: bernard.fortin@ecn.ulaval.ca

1 Introduction

Evaluating peer effects in academic achievement is important for parents, teachers and schools. These effects also play a prominent role in policy debates concerning ability grouping, racial integration and school vouchers. However, despite a growing literature on the subject (*e.g.,* Sacerdote 2001, Hanushek *et al.* 2003, Angrist and Lang 2004, Stinebrickner, R. and Stinebrickner 2006, Ammermueller and Pischke 2009), the evidence regarding the magnitude of peer effects on student achievement is mixed. This lack of consensus partly reflects various econometric issues that any empirical study on peer effects must address. Identifying and estimating peer effects raises at least three challenges. First, the relevant peer groups must be determined. Who interacts with whom? Second, peer effects must be identified from confounding factors. Especially, spurious correlation between students' outcomes may arise from self-selection into groups and from common unobserved shocks. Third, identifying the precise type of peer effect at work may be hard. The *reflection problem* may prevent separating contextual effects, *i.e.*, the influence of peers' characteristics, from endogenous effects, *i.e.*, the influence of peers' outcomes, see Manski (1993). Researchers have used various approaches to solve these three issues; we discuss the methods and results of previous studies in more detail in the next section. As will be clear, however, there is no simple methodological answer to these three challenges.

In this paper, we provide, to our knowledge, the first application of a novel approach developed by Lee (2007) for identifying and estimating peer effects. This approach is promising, as it allows to solve the problem of correlated effects and the reflection problem with standard data and without imposing *ad hoc* exclusion restrictions. The method does rely on structural assumptions, however, which makes its confrontation to real data particularly important. We have collected for this analysis original administrative data on test scores at the end of secondary school in the Canadian province of Quebec. We investigate the presence of peer effects in student achievement in mathematics, science, french, and history by applying Lee (2007)'s approach. In the process, we also advance our understanding of the method's properties.

Our econometric model relies on two key assumptions. First, individuals interact in groups that are known by the modeler. This means that the population of students is partitioned in groups (*e.g*., classes, grade levels) and that students are affected by all others in their groups but by none outside of it. This assumption is typical in studies of academic achievement but clearly arises from data constraints.[1] Second, individual outcome is determined by a linear-in-means model with group fixed effects. Thus, the test score of a student is affected

---

[1] Recent studies have looked at peer effects within finer social subdivisions such as networks of friendships among students, see Lin (2008) and Calvó-Armengol *et al.* (2009) for applications, and Bramoullé *et al.* (2009) and Lee *et al.* (2009) for theoretical analyses.

by his characteristics and by the average test score and characteristics in his peer group. In addition, it may be affected by any kind of correlated group-level unobservable. Lee (2007) shows that peer effects are fully identified in such a framework. The intuition for this result is the following. To address the problem of correlated effects, the group-level fixed effects must first be differentiated out. This removes any factor common to everyone in the group. In this model, however, a student is *not* assumed to be one of his own peer. This creates individual variations in average peer attributes, which are mechanically greater in smaller groups.[2] These variations survive the elimination of common unobservables. They lead to a non-linear impact of group sizes on reduced-form coefficients which can be used to recover the structural parameters.[3,4] More precisely, contextual *and* endogenous peer effects are identified when there are at least three groups of different sizes, see Davezies *et al.* (2009). When identified, the model can then be estimated through methods adapted from spatial econometrics, *e.g.* Cliff and Ord (1981).

We use administrative data on academic achievement for a large sample of secondary schools in the province of Quebec obtained from the Ministry of Education, Recreation and Sports (MERS). Our dependent variables are individual scores on four standardized tests taken in June 2005 (Math, Sciences, French and History) by fourth and fifth grade secondary school students. All 4th and 5th grade students in the province must pass these tests to graduate.[5] One advantage of these data is that all candidates in the province take the same exams, no matter their school and location. This feature effectively allows us to consider test scores as draws of a common underlying distribution. Another advantage is that our sample is representative and quite large. We have the scores of all students for a 75% random sample of Quebec schools which, over the four subjects, yields 194,553 test scores for 116,534 students. In terms of interaction patterns, the structure of the data leads us to make the following natural assumption. We assume that the peer group of a student contains all other students in the same school qualified to take the same test in June 2005. In practice, a small number of students postpone test-taking to August 2005. We extend Lee's methodology in the empirical modelling to address this issue. However, since the difference between observed group sizes and actual group sizes is small, the correction has little effect on the

---

[2]In contrast, these variations are absent in the model studied by Manski (1993). Manski studies a linear-in-expectations model which, in terms of identification, has the same properties as a linear-in-means model where the individual is included in his peer group. In this case, group fixed effects wash out both types of peer effects and identification fails to hold.

[3]This result is to be distinguished from the idea that the group size is a factor in a school's production function (*e.g.,* Angrist and Lavy 1999, Hoxby 2000b, Krueger 2003). In our model the effects of group sizes which are separable from the peer effects are controlled for by fixed effects in the structural model.

[4]This method also differs from the *variance contrast* approach developed by Graham (2008), which builds on Glaeser *et al.* (1996). The basic idea in this approach is that peer effects will induce intragroup dependencies in behavior that introduce variance restrictions on the error terms. These restrictions allow to identify the composite (endogenous + contextual) social interaction effects under the assumption that the variance matrix parameters are independent of the reference group size.

[5]These tests are taken once all courses in a subject matter are completed. Success to these tests only partly determine secondary school graduation, which also depends on performance in year-long evaluations.

results. Following Lee (2007), we estimate the model in two ways: through generalized two-stage least squares (G2SLS)[6] and, under stronger parametric conditions, through conditional maximum likelihood (CML).

Overall, we find that the CML method yields relatively precise estimates of peer effects.[7] In contrast, estimates from instrumental variables methods are quite imprecise with our data.[8] Based on CML estimates, the endogenous peer effect is positive, significant and quite high in Math (0.82) and History (0.65) but not significantly different from zero in French (0.33) and Science (−0.23).[9] These results are on the high side, but still within the spectrum, of estimates obtained in other studies. We also find evidence that contextual peer effects matter. For instance, interacting with older students usually has a negative effect on own test score independent of the impact of others' scores. While endogenous peer effects may capture direct emulation and imitation, contextual peer effects may reflect more indirect social interactions. In any case, this approach allows us to test typical restrictions used in the literature to get around the reflection problem. We often reject these restrictions in favor of the full, unrestricted model. This shows, *a posteriori*, the interest of disentangling the different types of peer effects.

One potentially important limitation of the method, however, is that it may not perform well when group sizes are large, see Lee (2007). More precisely, convergence in distribution of the peer effect estimates may occur at low rates when the number of groups is not much larger than average group size. Ratio between the two lies between 2.36 and 7.23 in our sample, so weak identification may be an issue. On the other hand, there is also much variance in group sizes in our samples. We suspect that this dispersion helps identification. We study this issue systematically through Monte-Carlo simulations. We find that indeed increasing group size variance has a strong positive impact on the precision of estimates. In addition, simulations calibrated on our data indicate that these small-sample problems are likely not an issue for our main empirical results.

The remainder of the paper is organized as follows. We discuss past research in section 2 and present our econometric model and the estimation methods in section 3. We describe our dataset in section 4. We present our empirical results in section 5 and run Monte Carlo experiments in section 6. We conclude in section 7.

---

[6]The instruments are directly derived from the structural model (Kelejian and Prucha 1998, Lee 2003, Lee 2007), see section 3.2.2 for details.

[7]The effect of individual characteristics, such as gender, age, and socioeconomic background, on test scores are precisely estimated by either method, and these estimates generally conform to expectations.

[8]The higher precision of CML estimates is consistent with results in Lee (2007) showing that CML estimators are asymptotically more efficient than IV estimators.

[9]For the French test, we also look at an alternative formulation under which the endogenous effect is equal to 0.51 and is significant.

2 Previous research

In this section, we give a brief overview of the recent literature on student achievement and peer effects, and we explain how our study complements and enhances current knowledge on peer interactions in academic outcomes.

As discussed above, measuring peer effects is complex as it raises three basic interrelated problems: the determination of reference groups, the problem of correlated effects and the reflection problem. The choice of reference groups is often severely constrained by the availability of data. In particular, there are still few databases providing information on the students' social networks; the Add Health dataset is an exception, see *e.g.* Lin (2008) and Calvo-Armengol *et al.* (2009).[10] For this reason, many studies focus on the grade-within-school level (*e.g.*, Hoxby 2000a, Hanushek *et al.* 2003, Angrist and Lang 2004).[11] Other studies analyze peer effects at the classroom level (*e.g.*, Kang 2007, Burke and Sass 2008, Atkinson *et al.* 2008, Ammermueller and Pischke 2009). Based on panel data covering Florida public school students in grades 3-10, Burke and Sass (2008) provide some evidence that peer effects tend to be weaker at the grade level than at the classroom level. This suggests two possible interpretations. On the one hand, learning and emulation effects between students are likely to be stronger at the classroom level. On the other hand, spurious correlated effects may also be more important at this level if the allocation of teachers and students to classes is not random (class-level selection biases). The administrative data we use in this study do not provide information on classes or teachers. Therefore, we assume that for each subject the relevant reference group for a student taking the test contains all other students in the same school who have completed all courses in the subject matter by June 2005. Thus, given that the reference group is likely to include students from other classes, one should probably expect peer effects to be smaller than at the classroom level.[12]

Two main strategies have been used to handle the problem of correlated effects. A first strategy has been to exploit data where students are randomly or quasi-randomly assigned within their groups (*e.g.*, Boozer and Cacciola 2001, Sacerdote 2001, Zimmerman 2003, Kang 2007). Results on the impact of contextual effects using randomly assigned roommates as peers are usually low though significant. Stinebrickner and Stinebrickner (2006) have argued that these studies tend to underestimate However, true peer effects as it is not clear that roommates represent peers of potential influence. Ammermueller and Pischke (2009) estimate peer effects for fourth grader in six European countries. They introduce school fixed effects to take into account nonrandom assignment

---

[10]Bramoullé *et al.* (2009) and Lee *et al.* (2009) determine conditions under which endogenous and contextual peer effects are identified when students interact through a social network known by the modeller and when correlated effects are fixed within subnetworks.

[11]These studies usually do not have information on teachers and classrooms.

[12]In fact, at the end of secondary level, classes and teachers are usually different depending on the subject matter taught.

of students across schools but they argue that there is no "teacher shopping" by parents or "student shopping" by teachers within schools. Therefore students are roughly randomly assigned to classes once the school is chosen and identification of contextual effects relies on variation across classes within schools. In our estimation approach, we also introduce school fixed effects to capture the feature that schools are not formed randomly. As in Ammermueller and Pischke' study, we find it plausible to assume that students in a given secondary school are randomly assigned to classes, when they reach grades 4 or 5.[13] Therefore, in our econometric approach we ignore the presence of a possible correlation between unobserved characteristics of teachers and of students which could induce sorting biases in our estimates. A second strategy exploits the availability of large panel administrative databases to introduce fixed individual, school, grade-within-school and cohort-by-grade effects (*e.g.*, Hoxby 2000a, Hanushek *et al.* 2003, Vigdor and Nechyba 2004, Carrell *et al.* 2008). In these studies, contextual effects are typically identified by the mobility of students into and out of the school and by changes in student circumstances over time (*e.g.*, parents' socioeconomic status). Of course, this approach cannot be used in our study since we do not have access to panel data.

The reflection problem is handled using three main strategies. In most papers, researchers estimate a reduced-form linear-in-means model, and no attempt is made to separate the contextual and endogenous peer effects. Only composite parameters are estimated (Sacerdote 2001, Ammermueller and Pischke 2009). In a second strategy, one uses instruments to obtain consistent estimates of the endogenous peer effect (*e.g.*, Evans *et al.* 1992, Gaviria and Raphael 2001, Atkinson *et al.* 2008). The problem here is to choose suitable instruments, that is, variables which are correlated with the endogenous peer effect but not correlated with the error terms in the structural model. For instance, Rivkin (2001) argues that the use of metropolitan-wide aggregate variables as instruments in the Evans *et al.* (1992) study exacerbates the biases in peer effect estimates. In our paper, we provide some results based on instrumental methods. However, our instruments are not *ad hoc* in the sense that they are derived from the structure of the model. A third strategy is to use lagged peer achievement as a proxy for current achievement (*e.g.*, Hanushek *et al.* 2003, Burke and Sass 2008). One problem is that this approach does not work when the error terms are autocorrelated. In any case, it requires panel data to be implemented.

In short, various strategies have been proposed to address the three basic issues that occur in the estimation of peer effects. But most rely on strong assumptions that are difficult to motivate and may not hold in practice. Some of them require panel data while others rely on experiments that randomly allocate students within their peer group. This makes the results in Lee (2007) particularly interesting, as they show that peer effects may be

---

[13]Informal discussions with secondary schools' administrators in the province of Quebec confirmed our intuition.

fully identified even with observational data in cross-section. This paper advances the literature on peer effects by providing the first application based on this new identification strategy.

3 Econometric model and estimation methods

3.1 Econometric model

We review and adapt the structural model suggested by Lee in the context of our application. Lee's model builds on and extends the standard linear-in-means model of peer effects (Moffitt 2001) to groups with various sizes. The set of students $\{i = 1, ...M\}$ is supposed to be partionned into groups of peers indexed by $r = 1, ..., R$. Let M_r be the rth group of peers, of size m_r. All students in the same group have the same number of peers since they interact with all others in the group. We assume that student i who belongs to group r is excluded from his own reference group. Let M_{ri} be student i's group of peers, of size $m_r - 1$.[14] A peer is any fellow student whose academic performance and personal characteristics may affect i's performance. Let y_{ri} be the test score obtained by student i. Let $\mathbf{x}_{ri}$ be a $1 \times K$ vector of characteristics of i and $\mathbf{X}_r$ be the $m_r \times K$ matrix of individual characteristics. For simplicity, the model is first presented with a unique characteristic ($K = 1$), which indicates whether the child comes from a low-income family. Another departure from the linear-in-means model is the inclusion of a term α_r that captures all group invariant unobserved variables (*e.g.*, same learning environment, similar preferences of school or motivation towards education). The error term ϵ_{ri} reflects other unobservable characteristics associated with i.

Importantly, we assume strict exogeneity of m_r and $\{x_{ri} : i = 1, ..., m_r\}$ *conditional* on the unobserved effect α_r, *i.e.*, $\mathbb{E}(\epsilon_{ri}|\ \mathbf{X_r}, m_r, \alpha_r) = 0$. This exogeneity assumption can notably accomodate situations where peer group size is endogenous. Suppose that, everything else equal, brighter students attend smaller schools, *i.e.*, schools where the cohort of students eligible to take the province-wide test in the subject matter (our peer groups) is small. In this case, peer group size m_r may well depend on unobserved common characteristics of the student's group, $\alpha_r : \mathbb{E}(\alpha_r|\ \mathbf{X_r}, m_r) \neq 0$. Our model allows for this type of correlation. However, conditional on these common characteristics, peer group size m_r is assumed to be independent of the student's idiosyncratic unobserved characteristics: $\mathbb{E}(\epsilon_{ri}|\ \mathbf{X_r}, m_r, \alpha_r) = 0$. We maintain this *anonymity* assumption throughout our analysis.

We do not change any other assumption of the linear-in-means model. In particular, we assume that a stu-

---

[14]We thus have $i \in M_r$ but $i \notin M_{ri}$.

6

6

dent's performance to the standardized test may be affected by the average performance in his group of reference, by his family socioeconomic background, and by the average socioeconomic background in his group. Formally, the structural model is given by:

$$y_{ri} = \alpha_r + \beta \frac{\sum_{j \in M_{ri}} y_{rj}}{m_r - 1} + \gamma x_{ri} + \delta \frac{\sum_{j \in M_{ri}} x_{rj}}{m_r - 1} + \epsilon_{ri}, \quad \mathbb{E}(\epsilon_{ri} | \, \mathbf{X_r}, m_r, \alpha_r) = 0, \tag{1}$$

where β captures the endogenous effect, γ the individual effect and δ the contextual effect. It is standard to require that $|\beta| < 1$. Except for this restriction (which is not imposed in our estimations and therefore could be tested), the model does not impose any other constraints on the parameters.[15]

One problem we face in our sample is that we do not always observe the scores of all students within a group. For instance, some students may postpone test-taking to the next session due to illness. We next extend our model to allow for this possibility. Our setting is one where the total number of students (including those who postpone test-taking) in each group is known, but we only observe the test scores of subsamples N_r of size n_r of each group M_r, with $n_r \le m_r$ and $\sum_{r=1}^{R} n_r = N$. We show how to adapt Lee's analysis to this more general setting. Let L_r be the complement of N_r, *i.e.*, $L_r = M_r - N_r$.[16] The structural equation becomes:

$$y_{ri} = \widetilde{\alpha}_r + \beta \frac{\sum_{j \in N_{ri}} y_{rj}}{m_r - 1} + \gamma x_{ri} + \delta \frac{\sum_{j \in N_{ri}} x_{rj}}{m_r - 1} + \epsilon_{ir}, \quad \mathbb{E}(\epsilon_{ri} | \, \mathbf{X_r}, m_r, n_r, \alpha_r) = 0, \tag{2}$$

where $\widetilde{\alpha}_r = \alpha_r + \beta \frac{\sum_{j \in L_r} y_{rj}}{m_r - 1} + \delta \frac{\sum_{j \in L_r} x_{rj}}{m_r - 1}$ is the new group fixed effect. Thus, even if we do not observe test scores for all students in each group, results are not necessarily biased due to a selection effect. The key observation is that in our model, effects stemming from unobserved individuals are the same for all students in each peer group. They are therefore picked up by the group fixed effect.

To eliminate group-invariant correlated effects, we next apply a *within* transformation to eq. (2). In particular, as we noted above, when the effect of group size is separable from peer and individual effects, it is captured by $\widetilde{\alpha}_r$. The model can address the problem of selection or endogenous peer group formation. For instance, school choice may depend on some unobserved factors specific to a school (*e.g.*, reputation, unobserved quality) and determine the type of students who are attracted by these schools. The advantage of the within transformation is that we compare students of the same type. This transformation also allows to control for common environment effects.

---

[15]This structural model can be derived from a choice-theoretic approach where each student's performance is obtained from the maximisation of his quadratic utility function which depends on his individual characteristics, his performance and his reference group's mean performance and mean characteristics. This approach also assumes that social interactions have reached a noncooperative (Nash) equilibrium at which expected performances are realized.

[16]If N_{ri} denotes the group of peers of student i, we also have $L_r = M_{ri} - N_{ri}$.

Resources available at the school level (*e.g.*, teaching, physical infrastructure) may affect the performance of all the students. Again, by comparing students within the same school, we can abstract from these effects. Importantly, mean peer behavior and mean peer characteristics vary within groups as, by assumption, student i is excluded from his reference group, that is, $i \notin N_{ri}$. In addition, these variations as mechanically smaller in larger groups, reflecting the diminishing impact of an additional individual on group means.[17] The within reduced form equation for students in the rth group is:

$$y_{ri} - \bar{y}_r = \frac{(m_r - 1)\gamma - \delta}{m_r - 1 + \beta}(x_{ri} - \bar{x}_r) + \frac{m_r - 1}{m_r - 1 + \beta}(\epsilon_{ri} - \bar{\epsilon}_r), \tag{3}$$

where means $\bar{y}_r$, $\bar{x}_r$ and $\bar{\epsilon}_r$ are computed over all observed students in the group. Note that only one composite parameter can be recovered from the reduced form for each group size m_r. At least three sizes are thus necessary to identify the three structural parameters β, γ and δ.

An intuitive interpretation for the identification of the endogenous peer effect from contextual peer effects is given in Bramoullé *et al.* (2009). In order to focus on the *reflection problem*, suppose that there are no correlated effects and that we observe the scores of all students within a group (for all r, $\widetilde{\alpha}_r = \alpha_0$). In the reduced form of this simpler model, own academic achievement only depends on own characteristics (*e.g.*, whether one is from a low-income background) and the fraction of peers from a low-income background, as follows:

$$y_{ri} = \frac{\alpha_0}{1 - \beta} + [\gamma + \frac{\beta(\gamma\beta + \delta)}{(1 - \beta)(m_r - 1 + \beta)}]x_{ri} + \frac{\gamma\beta + \delta}{(1 - \beta)(1 + \frac{\beta}{m_r - 1})}\bar{x}_i + \nu_i, \tag{4}$$

where ν_i is the error term[18] and where the proportion of peers from a low-income background $\bar{x}_i$ is now computed over all other observed students in the group, excluding student i. The effect of the parents' background x_{ri} may then be decomposed into a direct effect γ and an indirect effect that works through feedback effects: when j belongs to i's group, i's background affects j's academic achievement, which in turn, affects i's academic achievement.[19] The indirect effect decreases with group size m_r and becomes negligible as this size becomes larger. This reflects the diminishing role that i plays, by himself, in determining other students' academic achievement when the size of the group grows. For a similar reason, the reduced form coefficient on the fraction of low-income among i's peers is increasing in the size of the group. As the role played by any peer

---

[17]Formally, $|\frac{\sum_{j \in N_{ri}} x_{rj}}{m_r - 1} - \frac{\sum_{j \in N_{rk}} x_{rj}}{m_r - 1}| = \frac{1}{m_r - 1}|x_i - x_k|$ is decreasing with m_r.

[18]After some algebraic manipulations, one shows that: $\nu_i = [1 + \frac{\beta^2}{(1-\beta)(m_r - 1 + \beta)}]\epsilon_{ri} + \frac{\beta}{(1-\beta)(1 + \frac{\beta}{m_r - 1})}\bar{\epsilon}_i$.

[19]This indirect effect itself has different channels: i's background affects j's achievement both directly through a contextual effect, and indirectly through an endogenous effect (i.e., via the effect of i's achievement on j's achievement).

decreases, the effect of the fraction of peers from a low-income background becomes larger. Thus, variation in group sizes creates exogenous variation in the reduced form coefficients across groups which, in turn, yields identification. At the limit, as group sizes (the m_r's) become infinitely large, we cannot identify the endogenous peer effect from the contextual peer effect from the reduced-form equation (4). This suggests that the model may suffer from weak identification when interactions are structured in groups with large sizes.

3.2 Estimation methods

3.2.1 CML Estimator

We consider estimation under both strong (Conditional Maximum Likelihood or CML) and weaker (Instrumental Variables or IV) identification conditions.

To present CML and IV estimators, it is easier to express model (2) in matrix notations. We now allow for any number of characteristics, so that $\boldsymbol{\gamma}$ is a $K \times 1$ vector of individual effects and $\boldsymbol{\delta}$ a $K \times 1$ vector of contextual ones. Recall that in this setting, students are affected by all others in their group and by none outside of it. This means that the observed social interactions can be modelled as a $N \times N$ block-diagonal matrix $\mathbf{G} = Diag(\mathbf{G}_1, ..., \mathbf{G}_R)$, such that for all r, $\mathbf{G}_r$ is comprised of elements $g_{rij} = \frac{1}{m_r-1}$ if $i \neq j$ and $g_{rii} = 0$. In other terms, $\mathbf{G}_r = \frac{1}{m_r-1}(\boldsymbol{\iota}_{n_r}\boldsymbol{\iota}'_{n_r} - \mathbf{I}_{n_r})$, where $\boldsymbol{\iota}_{n_r}$ is a $n_r \times 1$ vector of ones and $\mathbf{I}_{n_r}$ the identity matrix of dimension n_r. Model (2) can be re-written in matrix form as follows:

$$\mathbf{y}_r = \boldsymbol{\iota}_{n_r}\widetilde{\alpha}_r + \beta\mathbf{G}_r\mathbf{y}_r + \mathbf{X}_r\boldsymbol{\gamma} + \mathbf{G}_r\mathbf{X}_r\boldsymbol{\delta} + \boldsymbol{\epsilon}_r, \tag{5}$$

where $\mathbb{E}(\boldsymbol{\epsilon}_r \mid \mathbf{X}_r, m_r, n_r, \alpha_r)=0$.

Applying the operator matrix $\mathbf{J}_r = \mathbf{I}_{n_r} - \frac{1}{n_r}\boldsymbol{\iota}_{n_r}\boldsymbol{\iota}'_{n_r}$ allows us to obtain deviations with respect to the mean for the observed group members. Pre-multiplying eq. (5) by $\mathbf{J}_r$ eliminates the group fixed effect and yields :

$$\mathbf{J}_r\mathbf{y}_r = \beta\mathbf{J}_r\mathbf{G}_r\mathbf{y}_r + \mathbf{J}_r\mathbf{X}_r\boldsymbol{\gamma} + \mathbf{J}_r\mathbf{G}_r\mathbf{X}_r\boldsymbol{\delta} + \mathbf{J}_r\boldsymbol{\epsilon}_r \tag{6}$$

Elementary linear algebra tells us that $\mathbf{J}_r\mathbf{G}_r = -\frac{1}{m_r-1}\mathbf{J}_r$. Letting $\mathbf{J}_r\mathbf{A}_r = \mathbf{A}_r^*$, we obtain

$$\frac{m_r - 1 + \beta}{m_r - 1}\mathbf{y}_r^* = \mathbf{X}_r^*\frac{(m_r - 1)\boldsymbol{\gamma} - \boldsymbol{\delta}}{m_r - 1} + \boldsymbol{\epsilon}_r^*. \tag{7}$$

which is equivalent to (3).

9

To derive the conditional maximum likelihood estimator, we assume that the ϵ_{ir}'s are i.i.d. $\mathbf{N}(0, \sigma^2)$. It follows that, given $\mathbf{X}_r$, m_r, and n_r, $\mathbf{y}_r^*$ follows a multivariate normal distribution with mean $\mathbf{X}_r^* \frac{(m_r-1)\boldsymbol{\gamma}-\boldsymbol{\delta}}{m_r-1+\beta}$ and variance $(\sigma \frac{m_r-1}{m_r-1+\beta})^2 \mathbf{J}_r$.[20] The log likelihood function can then be expressed as follows:

$$
\begin{aligned}
\ln L =\ & c + \sum_{r=1}^{R} (n_r - 1) \ln(m_r - 1 + \beta) - \frac{N-R}{2} \ln(\sigma^2) \\
& - \frac{1}{2\sigma^2} \sum_{r=1}^{R} \left(\frac{m_r-1+\beta}{m_r-1} \mathbf{y}_r^* - \mathbf{X}_r^* \frac{(m_r-1)\boldsymbol{\gamma}-\boldsymbol{\delta}}{m_r-1} \right)' \left(\frac{m_r-1+\beta}{m_r-1} \mathbf{y}_r^* - \mathbf{X}_r^* \frac{(m_r-1)\boldsymbol{\gamma}-\boldsymbol{\delta}}{m_r-1} \right),
\end{aligned}
$$

where c is a constant. This log likelihood function excludes any fixed effects. It is a *conditional* log likelihood function as it is conditional on the sufficient statistics $\bar{\mathbf{y}}_r$, (as well as on the $\mathbf{X}_r$'s, the m_r's, and the n_r's), for $r = 1, ... R$. Lee (2007) shows that the CML estimators of β, $\boldsymbol{\gamma}$, $\boldsymbol{\delta}$ and σ are consistent and asymptotically efficient under suitable regularity conditions and provided there is sufficient variation in group sizes.

3.2.2 2SLS and Generalized 2SLS estimators

Alternatively, the structural model (5) can be estimated by instrumental (IV) methods. To see how the methods work, define a $N \times N$ block-diagonal matrix $\mathbf{J} = Diag(\mathbf{J}_1, ..., \mathbf{J}_R)$. Concatenating eq. (6) over all groups yields:

$$
\mathbf{J}\mathbf{y} = \beta \mathbf{J}\mathbf{G}\mathbf{y} + \mathbf{J}\mathbf{X}\boldsymbol{\gamma} + \mathbf{J}\mathbf{G}\mathbf{X}\boldsymbol{\delta} + \mathbf{J}\boldsymbol{\epsilon}.
$$

where $\mathbf{y}$ (resp. $\mathbf{X}$) is obtained by stacking the vectors $\mathbf{y}_r$ (resp. the matrices $\mathbf{X}_r$), for $r = 1, ..., R$. The corresponding reduced-form is:

$$
\mathbf{J}\mathbf{y} = (\mathbf{I} - \beta \mathbf{G})^{-1} (\mathbf{J}\mathbf{X}\boldsymbol{\gamma} + \mathbf{J}\mathbf{G}\mathbf{X}\boldsymbol{\delta}) + (\mathbf{I} - \beta \mathbf{G})^{-1} \mathbf{J}\boldsymbol{\epsilon}.
$$

If $i \notin M_{ri}$ and there are at least three different group sizes, $\mathbb{E}[\mathbf{J}\mathbf{G}\mathbf{y}|\mathbf{X}, \mathbf{G}]$ is not perfectly collinear to $(\mathbf{J}\mathbf{X}, \mathbf{J}\mathbf{G}\mathbf{X})$ and the model is identified, see Bramoullé *et al.* (2009) for more details.[21] In particular, with exactly three different group sizes and one relevant characteristic (family socioeconomic background), $\mathbf{J}\mathbf{G}^2\mathbf{X}$ can be used as a valid instrument for $\mathbf{J}\mathbf{G}\mathbf{y}$ and the model is just identified.[22] When there are $K > 1$ relevant characteristics in model (5) and exactly three different group sizes, there are $2K + 1$ structural parameters and $3K$ reduced-form

---

[20]Note that only $n_r - 1$ elements of $\boldsymbol{\epsilon}_r^*$ are linearly independent.

[21]Remarkably, the usual case of group interactions and the more complex case of network interactions can be analyzed within the same formalism.

[22]In fact, $\mathbf{J}_r\mathbf{G}_r = -\frac{1}{m_r-1}\mathbf{J}_r$ and $\mathbf{J}_r\mathbf{G}_r^2 = \frac{1}{(m_r-1)^2}\mathbf{J}_r$, hence instruments are built here by premultiplying characteristics (in deviation) by group-dependent weights and by stacking them across groups.

parameters, so the model is over-identified. With $l > 3$ groups of different sizes and one relevant characteristic ($K = 1$), the model is also over-identified and the instrument set includes ($\mathbf{JG^2X}$, $\mathbf{JG^3X}$, ..., $\mathbf{JG}^{l-1}\mathbf{X}$).

One advantage of an IV approach over CML is that it requires less structure. Specifically, we do not assume that errors are homoskedastic (though we assume that they are independent *across* groups). Also we do not impose normality, nor use the structure on the error terms for identification purpose. Thus, identification in this case is semi-parametric, or "distribution-free".

In addition, we can obtain the "best" IV relatively easily from the structural model, see Lee (2007). The idea is to first obtain consistent estimates of the parameters and then to use these estimates to build the best possible instrumental variable. The procedure adapts Generalized 2SLS strategies proposed in Kelejian and Prucha (1998) and refined in Lee (2003) to our current setting. It yields an asymptotically optimal estimator in the class of IV estimators. More precisely, our first step consists in estimating a 2SLS as described above, by using as instruments

$$\mathbf{S} = \left[\begin{array}{ccc} \mathbf{JX} & \mathbf{JGX} & \mathbf{JG^2X} \end{array} \right],$$

At this step, the model is overidentified, and we obtain $\widehat{\boldsymbol{\theta}}^{2SLS} = (\widetilde{\mathbf{X}}'\mathbf{P}\widetilde{\mathbf{X}})^{-1}\widetilde{\mathbf{X}}'\mathbf{Py^*}$, where

$$\widetilde{\mathbf{X}} = \left[\begin{array}{ccc} \mathbf{JGy} & \mathbf{JX} & \mathbf{JGX} \end{array} \right]$$

is the matrix of explanatory variables, $\mathbf{P} = \mathbf{S}(\mathbf{S}'\mathbf{S})^{-1}\mathbf{S}'$ is the weighting matrix and, recall, $\mathbf{y}^* = \mathbf{Jy}$. Then, define $\boldsymbol{\theta} = \left[\begin{array}{ccc} \boldsymbol{\beta} & \boldsymbol{\gamma} & \boldsymbol{\delta} \end{array} \right]$. The second step consists in estimating a 2SLS estimator using as instruments $\widehat{\mathbf{Z}} = \mathbf{Z}(\widehat{\boldsymbol{\theta}}^{2SLS})$, with

$$\mathbf{Z}(\boldsymbol{\theta}) = \left[\begin{array}{ccc} \mathbb{E}[\mathbf{JGy}(\boldsymbol{\theta}) \mid \mathbf{X}, \mathbf{G}] & \mathbf{JX} & \mathbf{JGX} \end{array} \right].$$

From the reduced-form equation, it follows that

$$\mathbb{E}[\mathbf{JGy}(\boldsymbol{\theta}) \mid \mathbf{X}, \mathbf{G}] = \mathbf{G}(\mathbf{I} - \beta\mathbf{G})^{-1}[\mathbf{J}(\mathbf{X}\boldsymbol{\gamma} + \mathbf{GX}\boldsymbol{\delta})].$$

At this step, the model is just identified, and we obtain

$$\widehat{\boldsymbol{\theta}}^{G2SLS} = (\widehat{\mathbf{Z}}'\widetilde{\mathbf{X}})^{-1}\widehat{\mathbf{Z}}'\mathbf{y}^*.$$

The variance matrix of the estimated parameters is consistently estimated by :

$$\widehat{V}(\widehat{\boldsymbol{\theta}}^{G2SLS}) = (\widehat{\mathbf{Z}}'\widetilde{\mathbf{X}})^{-1}\widehat{\mathbf{Z}}'\mathbf{D}\widehat{\mathbf{Z}}(\widetilde{\mathbf{X}}'\widehat{\mathbf{Z}})^{-1},$$

where $\mathbf{D}$ is an $N \times N$ block-diagonal matrix with entries in each block r given by the products of the residuals from this second step.

4   Data

We gathered for this analysis original data from the Quebec Government MERS. These administrative data provide detailed information on individual scores on standardized tests taken in June 2005 on four subjects (Math, Sciences, French and History) by fourth and fifth grade secondary school students. They also include information on the age, gender, language spoken at home and socioeconomic status of students. Sampling has been done in two steps. The population of interest is the set of all fourth and fifth grade secondary school students who are candidates to the MERS examinations in June 2005. This population is comprised of 152,580 students in total. In the first step, a 75% random sample of secondary schools offering fourth and fifth grade classes in the 2004-2005 school year have been selected. In the second step, all fourth and fifth grade students in these schools have been included. Overall, we have 194,553 individual test scores for 116,534 students.[23]

One limitation of these data is that we can only group students into grade-level within each school. On the other hand, there are many advantages to its use. First, all 4th and 5th grade students must take these tests to qualify for secondary school graduation. This means that our results do not pertain to a selected sample of schools. In particular, both public and private school students have to take these tests. Another advantage is that the tests are standardized, *i.e.*, designed and applied uniformly within the province of Quebec. We use test results gathered by the MERS, so there is less scope for measurement error with these data than with survey data on grades. Finally, although survey data may have provided information on a larger set of covariates, sample sizes in our study are larger than in typical school surveys.

Notwithstanding the limited information on the group of interaction, we know that in order to qualify for a test a student should complete all coursework in the subject matter. This helps us to limit peer groups to the population targeted by the tests. More specifically, we assume that the relevant reference group for a student taking a test is comprised of all other students in the same school who are qualified to take the test in June 2005.

---

[23]A student can take more than one subject test.

Two test sessions are offered for those who completed coursework in the Spring semester. We thus consider as belonging to the same group all those who belong to the same school and who take a subject test in one of the two consecutive sessions of June and August 2005. We know the number of students in each of these groups. But we only observe test scores for the set of students who took the test in June. Therefore we do not always observe the scores of all students within a group. We offered a correction for this problem in our discussion of the econometric model, and our empirical results below incorporate this correction. In any case an overwhelming majority of the students do take the tests in June, so the correction has little effect on the results.

We use for this study French, History, Science and Math test results as reported in the MERS administrative data. Students in a regular track take History and Science tests in Secondary 4. The French test is commonly taken in Secondary 5. For this test only, and because of the obvious impact of language spoken at home, we also look at the subsample of students whose mother tongue and language spoken at home are the same as the language of instruction (French2).[24] Finally, we focus on students who take the Math test in Secondary 5 (Math 514). This completes their mathematical training for secondary school.[25] We focus on this test in our analysis.

We provide descriptive statistics in Table 1. For each subject, the dependent variable in our econometric model is the test score obtained in the provincial standardized test. The average score is between 70% and 75% in French, Science and History tests. It is lower and about 62% in Math. In samples for which the regular track for the test is Secondary 5 (resp. Secondary 4), the average age of students is close to 16 (resp. 15). Most students taking French and Math (98% and 96%) are enrolled in Secondary 5. Most of those taking Science and History are enrolled in Secondary 4 (92% and 96%). Between 52% and 55% of students are female, and between 11% and 13% of students speak a language at home which is different from the language of instruction.[26] Between 30 and 34% of students come from a relatively high socioeconomic background and between 40% and 42% from a medium one.[27]

We observe test scores and characteristics of students taking the same test in June 2005. Sample sizes are 41, 778 for French, 54, 981 for Science, 15, 771 for Math, and 55, 057 for History. When we exclude students

---

[24]Thus, models estimated for this subsample assume that the interaction group is comprised of all other students in the same school qualified to take the test in June 2005 *and* whose mother tongue and language spoken at home is similar to the language of instruction.

[25]The MERS administers a unique test to all secondary school students in French, History and Science. In contrast, it administers different tests in Math, depending on academic options chosen early on by the students. We report here results for students following the regular mathematical training (Math 514).

[26]The language of instruction is French in most schools, and English otherwise.

[27]We use an index of socio-economic status provided by the MERS. This index is computed from data from the 2001 census. It uses information on the level of education of the mother (a weight of 2/3) and the job status of parents (weight of 1/3). Low socio-economic status corresponds to the three lowest deciles of the index (high socio-economic status to the three highest deciles).

whose mother tongue or language spoken at home is different from the language of instruction, the number of observations for the French sample drops to $37,143$. We also observe the number of students who completed coursework but postpone test-taking to August 2005. There are 118 (89) students postponing French (French2), 186 postponing History, 195 postponing Science, and 160 postponing Math. We observe between 307 and 382 peer groups depending on the subject matter considered. The average group size is between 50 (Math) and 146 (Science). The ratio between the number of groups and the average group size varies between 2.36 (French) and 7.23 (Math). These numbers are relatively small, which suggests that our estimates could be subject to weak identification problems. The group size standard deviation is quite large, however, varying between 50 (in Math) and about 135 (in Science and History). We expect such dispersion in group sizes to help identification. We analyze these issues in more details in Section 6.

5 Empirical Results

5.1 Naive estimates

We first report naive OLS estimates in Table 2. These estimates provide a useful benchmark with which to compare the results from our more sophisticated methods. We ignore here the two key econometric issues of correlated unobservables and simultaneity, and simply regress individual test score on individual characteristics, average score of peers and average characteristics of peers. So the OLS estimator is not consistent even under the null of no correlated effects. The estimated impact of others' test scores on one own test score is quite high (about 0.8) and statistically significant for all subjects. Also, most individual and contextual variables are significant. Interestingly, individual and contextual effects corresponding to the same characteristic often have opposite signs. We will see below that many of these features are not robust to addressing the main econometric issues.

5.2 CML estimates

We next report the results of CML estimation in Table 3. The model estimated is the linear-in-means model with group fixed effects, individual impacts, and endogenous and contextual peer effects. We find that the estimated endogenous peer effect lies between -0.24 and 0.83. It is significantly different from zero and positive for French2 ($\widehat{\beta} = 0.51$), Math ($\widehat{\beta} = 0.82$), and History ($\widehat{\beta} = 0.65$). It is not significant for French ($\widehat{\beta} = 0.33$) and for Science ($\widehat{\beta} = -0.23$). Note that for all subjects the endogenous effect is smaller that the one obtained using a naive OLS. This is likely to reflect a positive correlation between unobservables and test scores. Such positive correlation could emerge, for instance, from variations in teacher quality.

14

How do these results compare with those obtained in other studies? As a number of other researchers (*e.g.*, Carrell *et al.* 2008), we find that the endogenous peer effect on achievement is the largest in Math. This may reflect the fact that Mathematics provide more opportunities for interactions among students. In Math, Hoxby (2000a) reports a 0.1 to 0.55-point increase in own score in relation with a 1-point increase in mean score of peers. Hanushek *et al.* (2003) show a 0.4-point increase in association with a 1-point increase in mean math score of peers. Zimmer and Toma (2000) report an endogenous math peer effect lying between 0.6 and 0.8. On the lowest side of the spectrum, Vigdor and Nechyba (2004) obtain an endogenous math peer effect of 0.07.[28] So our estimate lies on the high side of the range of previous estimates.

Most of the individual characteristics have a significant effect on test scores, and the signs of these effects essentially conform to expectations. All test scores decrease significantly with age. Since older students have often repeated a grade, being younger is a natural proxy for ability. Test scores are significantly higher for female students than for male students, except for History where male students perform significantly better than female students. This is broadly consistent with results from previous studies.[29] The performance of foreign students is, non surprisingly, significantly lower than for non-foreign students on the French test, but higher for Science and History and not significantly different for Math. Secondary 5 students tend to perform significantly better on all tests than Secondary 4 students, which reflects the positive impact of an additional year of schooling on test scores. Finally, students from a higher socioeconomic category perform significantly better in all tests. Interestingly, estimates of individual effects are close to the ones obtained from naive OLS regressions.

However, this is not the case as far as contextual variables are concerned. A few of these variables have a significant impact on student performance. Average age of other students has a negative and significant effect on all test scores except Math where it is positive but not significant. Proportion of other students enrolled in Secondary 5 have a large positive and significant effect on own score in French. Peers' socioeconomic background has little effect on own schooling performance. When significant, the magnitude of contextual effects is always larger than the magnitude of individual effects. This is not surprising as it captures the effect of a unit change of the characteristic of *every* other student in the group. Contrary to the OLS estimates, when significant, contextual effects are in general of the same sign as the impact of the individual characteristic to which they are associated.[30] This is consistent, for instance, with situations where students work together and may pick up

---

[28]Kang (2007, p. 475) provide a more complete survey of endogenous peer effects in achievement in mathematics.

[29]For instance, results from the 2000 Program for International Student Assessment (PISA) show that Quebec female students perform better than males on reading literacy tests but that the differences in performance on mathematics and science tests are smaller and not significant, see Quebec Government (2001). Similarly, in our analysis, the difference in performance is quantitatively large in French but much smaller in the other disciplines.

[30]The only exception is the proportion of other female students which has a negative effect on own performance in the French test.

work habits of others. Thus, if older students exert less effort, being in a group with a higher proportion of older students may lead to work less, and this may have a depressing effect on grades independent of the other peer effects.

5.3 Reflection problem

We next take advantage of our framework to examine common ways of getting around the simultaneity problem that do not rely on variations in group sizes. One such way is simply to impose that $\beta = 0$ (no endogenous effect). Thus, we estimate a model with individual and contextual effects only. In this case the CML estimator is equivalent to a fixed effects OLS estimator. We report results in Table 4. We find that estimated coefficients of individual characteristics are similar to those of the full CML model. However, estimates of contextual effects are quite different. For instance, while a rise in own age is still associated with a decline in own test score, peers' average age is now found to have a positive and significant effect on own test score for all tests.[31] Of course, we know from the CML estimation of the full model that the endogenous effect is significantly different from zero for most subjects. Thus, imposing $\beta = 0$ is not appropriate here, and the positive age contextual effect is likely picking up the omitted endogenous effect. Note, however, that we need the results from the full estimation to reach this conclusion.

Another possible way of addressing the simultaneity problem without exploiting group size variations is to exclude at least one contextual variable from the outcome equation and to use it as an instrument for average test score. We estimate a model similar to the one presented in Table 3 but with no contextual effects (*i.e.*, imposing $\boldsymbol{\delta} = \mathbf{0}$); see Table 5. Using likelihood ratio tests, we reject the null of no contextual effects for History and the two French samples but not for Science and Math. This suggests that the exclusion restrictions may be valid for these other samples. Therefore, the ML estimators provided in Table 5 should be consistent and asymptotically more efficient than those provided in Table 2 for the Science and Math tests. Observe again, however, that we could not have known this *a priori* without an estimation of the full model.

Overall, this shows the interest of Lee's solution to the reflection problem. Estimating a model with both endogenous and contextual peer effects is needed to recover precisely the different types of peer effects at work.

---

[31]Note that the results presented in Table 4 cannot be interpreted as the coefficients of the reduced form of the model with endogenous effect. Because of the nonlinear effect of group size on reduced form coefficients, β is identified and the reduced forms for cases where $\beta = 0$ and $\beta \neq 0$ differ. This stands in contrast to the analysis in Manski (1993).

5.4 2SLS and G2SLS estimates

We next contrast IV and CML estimates. We present in Tables 6 and 7 the 2SLS and G2SLS estimation results of the linear-in-means model of peer effects with group fixed effects, individual impacts, and endogenous and contextual peer effects. In contrast to the CML estimates of Table 3, none of the endogenous effects is now statistically significant. This is consistent with Lee's (2007, p. 345) result that the asymptotic efficiency of IV estimators is smaller than that of the CML. Estimated individual effects are quite similar to the corresponding CML estimates. Some contextual effects are similar while others are different. For instance, the proportion of other students in Secondary 5 still has a large and positive effect on own French score as well as no significant effects for the other subjects. In contrast, average age among peers now has a positive and significant effect on own score for most subjects, rather than a negative one. This could be explained by differences in small sample properties of both methods, possibly aggravated by the imprecision in the estimation of the endogenous peer effect.

Table 6 also reports two standard test results giving information on instrumental variables properties. We first look at Sargan tests on the validity of the instruments and the over-identification restrictions of the model. We do not reject the null for Science, Math and History, but we reject it for French (and French2). While this may indicate a problem of model specification in these two last cases, one must be cautious in interpreting the test given the likely low convergence of peer effects IV estimates. We then compute Stock and Yogo (2005)'s test statistics on weak identification.[32] Based on the definition that a group of instruments is weak when the bias of the IV estimator relative to the bias of ordinary least squares exceeds a certain threshold b, say 5%, one rejects the null that the instruments are weak for all subject matters.

6 Monte Carlo simulations

In this section, we study through simulations the effect of group sizes and their distribution on the precision and bias of our estimates. Lee (2007) shows that the CML and IV estimators may converge in distribution at low rates when the ratio between the number of groups and the average group size is small. Since this ratio varies between 2.36 and 7.23 in our samples, a problem of weak identification could in principle emerge. However, the standard deviation of the distribution of group sizes is also relatively large, and we suspect that this helps identification. To study these issues, we realize two simulation exercices. First, we vary group sizes in a systematic manner

---

[32]The Stock and Yogo test assumes that the instruments are valid, so it may be biased in the case of French and French2.

and study how this affects the bias and precision of CML and IV estimators. We notably vary the group size standard deviation and partly calibrate simulation parameters on our data. Second, we look at bias and precision of estimates for *fully* calibrated simulations, when group sizes are exactly the same as in the data. Overall, while our analysis confirms Lee's earlier results, we also find a strong positive impact of the standard deviation of group sizes on the strength of identification. Especially, conditional maximum likelihood performs well on fully calibrated simulations. This indicates that the bias due to small sample issues is likely low in the results presented in Table 3.

For each simulation exercise, we keep the number of observations fixed around $42,000$, and run $1,000$ replications. We first consider average sizes of 10, 20, 40, 80 and 120. We pick group sizes from the following intervals with decreasing standard deviation:

- Average size of 10 : $[3, 17]$, $[5, 15]$, $[7, 13]$ and $[9, 11]$,

- Average size of 20 : $[3, 37]$, $[8, 32]$, $[13, 27]$ and $[18, 22]$,

- Average size of 40 : $[3, 77]$, $[12, 68]$, $[21, 59]$, $[30, 50]$ and $[39, 41]$,

- Average size of 80 : $[3, 157]$, $[18, 142]$, $[33, 127]$, $[48, 112]$ and $[63, 97]$,

- Average size of 120 : $[3, 237]$, $[28, 212]$, $[53, 187]$, $[78, 162]$ and $[103, 137]$.

For each of the intervals described above, we proceed in the following manner:

1. pick a group size from a uniform distribution for which the support is defined by the minimum and maximum value of the interval;

2. truncate this value by eliminating its decimal portion;

3. repeat step 1 and 2 as long as the total number of observations is below or equal $42,000$.

To reduce computing time, we assume that all students have the same characteristics except for age and gender. We assume that age follows a normal distribution and gender follows a Bernoulli distribution. We calibrate the moments of these distributions on the sample of students taking the French test: average age is 16, variance of age is 0.25, and proportion of girls is 0.55. Values of the structural parameters β, γ and δ are set to

the estimated coefficients for the French test:

$$\beta = 0.35,$$

$$\gamma_{age} = -8, \ \gamma_{gender} = 3.8,$$

$$\delta_{age} = -40, \ \delta_{gender} = -25.$$

We assume that the values of ϵ in the structural equation are drawn randomly from a normal distribution with mean zero and variance $\sigma^2 = 1$. We generate the endogenous variable y from the reduced-form equation in deviation form. We estimate the model using CML and IV (G2SLS).

Looking at Table 8, we first compare simulation results across average group sizes and then we examine how estimators perform for a given average group size as variance in group size decreases. Separate horizontal panels in Table 8 pertain to different values of average group size. We report the average estimated coefficient and standard error by CML and G2SLS for the endogenous effect (first vertical panel), the exogenous effect associated with age (second vertical panel) and the exogenous effect associated with gender (third vertical panel). We find that even for the largest average group size (*i.e.*, 120), both CML and IV perform well in terms of bias and precision (first line in the last horizontal panel of Table 8). The biases of CML and IV get in general larger as average group size increases. The CML estimate of the endogenous effect attains a plateau at the value 1, while the IV estimate becomes rapidly larger (in absolute value) than 1. This is consistent with the fact that the CML estimator tends towards the naive OLS estimator as group sizes become larger. In contrast, the bias on the endogenous effect estimated by IV rapidly exceeds the bias of OLS. This is consistent with instruments becoming less correlated to peer average outcome as group sizes increase. In general, peer effects are also less precisely estimated in large groups than in small groups, regardless of the estimation method. But, for a given Monte Carlo experiment, the magnitude of the bias and the loss in precision are always larger for IV than for CML.

Our main new result concerns the effect of group size dispersion. When we fix the value of the average group size and reduce the length of the interval from which group sizes are picked, we find that the bias of both CML and IV typically increases while the precision typically decreases. In Table 8, this amounts to looking at each horizontal panel separately. Since we roughly pick group sizes from a uniform distribution holding average group size fixed, reducing the interval's length corresponds to a reduction in variance. Not surprisingly, bias of IV rises more rapidly than bias of CML when the variance in group sizes is reduced, and the precision declines more rapidly as well.

19

We next fully calibrate the simulations' parameters on the data, We use observed group sizes in the French sample, calibrate the model parameters $\{\beta, \gamma_{age}, \gamma_{gender}, \delta_{age}, \delta_{gender}\}$ and moments of the explanatory variables as previously, and set the variance of the error term in the structural model equal to the estimated variance in the French sample ($\hat{\sigma}^2 = 154.7$). Simulation results are reported in Table 9. The CML estimator has small bias and standard error, while the IV estimator is not precisely estimated and the bias is large. These results confirm what we obtained from picking group sizes at random; they show that dispersion in group sizes help identification. Overall, this suggests that small sample bias may be relatively high in the IV estimates of Tables 6 and 7 but relatively low for the CML estimates of Table 3.

7 Conclusion

It is now well documented that identification of peer effects on student achievement is challenging because of the difficulties to isolate them from other confounding effects. A number of studies have addressed this problem by exploiting data where students are randomly assigned to groups or by imposing *ad hoc* exclusion restrictions on the structural model determining student achievement. In this paper, we adopt a different identification strategy. We provide the first empirical application of Lee (2007)'s novel approach, which shows that endogenous and contextual peer effects can be identified when agents interact in groups and when there are sufficient variations in group sizes. Correlated unobservables are allowed as long as they are fixed within each group. The model is applied to original administrative data providing individual scores on standardized tests taken in June 2005 in four subjects by fourth and fifth grade secondary school students in the province of Quebec (Canada). The results generally indicate that students benefit from their peers' higher test scores. Thus, based on our conditional maximum likelihood results, a point increase in the average test score of his peers increases a student's test score by 0.5 in French, 0.65 in History and 0.83 in Math. Contextual peer effects also matter. For instance, interacting with older students usually has a negative and significant effect on scores. Overall, our analysis confirm the interest of this new methodology to study peer effects with cross-section observational data. From a practical standpoint, its main limitation is that it needs enough variation in group sizes to perform well. Our Monte-Carlo simulations and empirical results show that the amount of variation present in our data is likely sufficient to obtain reasonable estimates, thank to the typically high variance in school sizes.

Our research could be extended in many directions. It would be interesting to evaluate the validity of this approach by using data where group membership is experimentally manipulated *and* group sizes are heterogenous (as in Sacerdote 2001). Such data would also allow to test the validity of the structural model. It would also be

interesting to analyze how group size variations may help to identify peer effects when the outcome is a discrete variable (*e.g.,* pass or fail). Brock and Durlauf (2001, 2007) have studied peer effects identification with discrete outcomes but they ignore group size variations. A third potentially fruitful direction of research would be to analyze and estimate a nonlinear version of Lee's approach. For instance, student achievement could depend on the mean *and* standard deviation of peers' attributes. Overall, we think that this first empirical application confirmed the interest of the method. Many more applications in different settings are needed, however, in order to gain a thorough understanding of the method's advantages, limitations, and applicability.

REFERENCES

Ammermueller, A. and Pischke, J-S. (2009): "Peer Effects in European Primary Schools: Evidence from the Progress in International Reading Literacy Study", *Journal of Labor Economics,* Vol. 27(3), 315-348.

Angrist, J. D. and Lang, K (2004): "Does School Integration Generate Peer Effects? Evidence from Boston's Metco Program", *American Economic Review,* Vol. 94(5), 1613-1634.

Angrist, J. D. and Lavy, V. (1999): "Using Maimonides' Rule to Estimate the Effect of Class Size on Student Achievement", *Quarterly Journal of Economics,* Vol. 114(2), 533-575.

Atkinson, A., Burgess, S., Gregg, P., Propper, C. and Proud, S. (2008): "The Impact of Classroom Peer Groups on Pupil GCSE Results", WP no. 08/187, Center for Market and Public Organisation, Bristol, UK.

Boozer M. A. and Cacciola, S. E. (2001): "Inside the "Black Box "of Project STAR: Estimation of Peer Effects Using Experimental Data", WP no. 832, Economic Growth Center, Yale University.

Bramoullé, Y., Djebbari, H. and Fortin, B. (2009): "Identification of Peer Effects Through Social Networks", *Journal of Econometrics,* Vol. 150, 41-55.

Brock, W. and Durlauf, S. (2001): "Discrete Choice with Social Interactions", *Review of Economic Studies,* Vol. 68, 235-260.

Brock, W. and Durlauf, S. (2007): "Identification of Binary Choice Models with Social Interactions". *Journal of Econometrics,* Vol. 140, 57-75.

Burke, M.A. and Sass, T.R. (2008): "Classroom Peer Effects and Student Achievement", WP no. 18, National Center for Analysis of Longitudinal Data in Education Research.

Calvó-Armengol, A., Patacchini, E. and Zenou, Y. (2009): "Peer Effects and Social Networks in Education", *Review of Economic Studies*, Vol. 76(4), 1239-1267.

Carrel, S.E., Fullerton, R.L. and West. J.E. (2008): "Does Your Cohort Matter? Measuring Peer Effects in College Achievement",WP no. 14032, National Bureau of Economic Research, Cambridge, MA.

Cliff, A. and Ord J. K. (1981): *Spatial Processes.* London: Pion.

Davezies, L., d'Haultfoeuille, X. and Fougère, D. (2009): "Identification of Peer Effects Using Group Size Variation", *Econometric Journal*, Vol. 12, 397-413.

Evans, W., Oates, W. and Schwab, R. (1992): "Measuring Peer Group Effects: a Study of Teenage Behavior", *Journal of Political Economy,* Vol. 100(5), 966-991.

Gaviria, A. and Raphael, S. (2001): "School based Peer Effects and Juvenile Behavior", *Review of Economics and Statistics*. Vol. 83(2), 257-268.

Glaeser, E., Sacerdote, B. and Scheinkman, J. (1996): "Crime and Social Interaction", *Quarterly Journal of Economics,* Vol. CXI, 507-548.

Graham, B.S. (2008): "Identifying social interactions through conditional variance restrictions", *Econometrica,* Vol. 76 (3), 643-660.

Hanushek, E., Kain, J., Markman, J. and Rivkin, S. (2003): "Does Peer Ability Affect Student Achievement?", Vol. 18(5), *Journal of Applied Econometrics,* 527-544.

Hoxby, C. (2000a): "Peer Effects in the Classroom: Learning from Gender and Race Variation", WP no. 7867, National Bureau of Economic Research, Cambridge, MA.

Hoxby, C. (2000b): "The Effects of Class Size on Student Achievement: New Evidence from Population Variation", *Quarterly Journal of Economics,* Vol. 115(4), 1239-1295.

Kang, C. (2007): "Classroom Peer Effects and Academic Achievement: Quasi-Randomization Evidence from South Korea", Vol. 61*, Journal of Urban Economics,* Vol. 61, 458-495.

Kelejian H.H. and Prucha I.R. (1998): "A Generalized Spatial Two-Stage Least Squares Procedure for Estimating a Spatial Autoregressive Model with Autoregressive Disturbances", *Journal of Real Estate Finance and Economics,* Vol 17*,* 99-121.

Krueger, A.B. (2003): "Economic Considerations and Class Size", *Economic Journal,* Vol. 113(485), F34-F63.

Lee, L. F. (2003): "Best Spatial Two-Stage Least Squares Estimators for a Spatial Autoregressive Model with Autoregressive Disturbances", *Econometric Reviews,* Vol. 22(4), 307-335.

Lee, L. F. (2007): "Identification and Estimation of Econometric Models with Group Interactions, Contextual Factors and Fixed Effects", *Journal of Econometrics,* Vol. 140(2)*,* 333-374*.*

Lee, L. F., Liu, X. and Lin X. (2009): "Specification and Estimation of Social Interaction Models with Network Structure, Contextual Factors, Correlation and Fixed Effects", *Econometric Journal*, forthcoming.

Lin, X. (2008): "Identifying Peer Effects in Student Academic Achievement by Spatial Autoregressive Models with Group Unobservables", *Mimeo*, Department of Economics, Tsinghyua University, Beijing.

Manski, C. (1993): "Identification of Endogenous Social Effects: The Reflection Problem", *Review of Economic Studies*, Vol. 60(3), 531-542.

Moffitt, R. (2001): "Policy Interventions, Low-Level Equilibria, and Social Interactions", in *Social Dynamics,* edited by Steven Durlauf and Peyton Young, MIT press.

Quebec Government. (2001): "PISA 2000. The Performance of Canadian Youth in Reading, Mathematics and Science. Results for Quebec Students Aged 15", Ministry of Education, Recreation and Sports.

Rivkin, S. G. (2001): "Tiebout Sorting, Aggregation and the Estimation of Peer Group Effects", *Economics of Education Review,* Vol. 20, 201-209.

Sacerdote, B. (2001): "Peer Effects with Random Assignment: Results for Darmouth Roommates", *Quarterly Journal of Economics*, Vol. 116(2), 681-704.

Stinebrickner, R. and Stinebrickner, T. R. (2006): "What Can be Learned About Peer Effects Using College Roommates? Evidence from New Survey Data and Students from Disadvantaged Backgrounds", *Journal of Public Economics,* Vol. 90, 1435-1454.

Stock, J. H. and Yogo, M. (2005), "Testing for Weak Instruments in Linear IV Regression.", in *Identification and Inference for Econometric Models: Essays in Honor of Thomas J. Rothenberg,* edited by D. W. K. Andrews and J. H. Stock, Cambridge University Press.

Vigdor, J. L. and Nechyba, T. (2004): "Peer Effects in Elementary Schools: Learning from "apparent"random assignment.", *Mimeo,* Department of Economics, Duke University.

Zimmer, R. W. and Toma, E. (2000): "Peer Effects in Private and Public Schools Across Countries", *Journal of Policy Analysis and Management,* Vol. 19, 75-92.

Zimmerman, D. (2003): "Peer Effects in Academic Outcomes: Evidence from a Natural Experiment", *Review of Economics and Statistics,* Vol. 85(1), 9-23.

Table 1: Descriptive statistics

Course	Variable	Mean	S.D.
French	Score	72.647	14.086
(Sec. 5)	Age	16.142	0.488
	Socio-ec. Index	-	-
	Perc. High	0.328	0.469
	Perc. Med.	0.409	0.492
	Gender (Female=1)	0.549	0.500
	Foreign	0.111	0.310
	Secondary 5	0.985	0.120
	Number of observations	41778	
	Number of groups	314	
	Avg. size of true groups	133.4	115.7
	Avg. size of observed groups	133.1	115.4
French2 †	Score	73.102	13.836
(Sec. 5)	Age	16.118	0.450
	Socio-ec. Index	-	-
	Perc. High	0.333	0.470
	Perc. Med.	0.417	0.493
	Gender (Female=1)	0.551	0.497
	Foreign	-	-
	Secondary 5	0.986	0.120
	Number of observations	37143	
	Number of groups	307	
	Avg. size of true groups	121.3	112.3
	Avg. size of observed groups	121	112.1
Science	Score	74.689	17.671
(Sec. 4)	Age	15.255	0.610
	Socio-ec. Index	-	-
	Perc. High	0.338	0.470
	Perc. Med.	0.402	0.490
	Gender (Female=1)	0.527	0.499
	Foreign	0.127	0.333
	Secondary 5	0.077	0.267
	Number of observations	54981	
	Number of groups	378	
	Avg. size of true groups	146.0	134.2
	Avg. size of observed groups	145.5	133.7

Table 1: Descriptive statistics (continued)

Course	Variable	Mean	S.D.
Math ††	Score	62.088	15.83
(Sec. 5)	Age	16.272	0.574
	Socio-ec. Index	-	-
	Perc. High	0.303	0.460
	Perc. Med.	0.400	0.490
	Gender (Female=1)	0.540	0.498
	Foreign	0.111	0.314
	Secondary 5	0.957	0.202
	Number of observations	15771	
	Number of groups	361	
	Avg. size of true groups	50.7	49.9
	Avg. size of observed groups	49.9	49.7
History	Score	70.156	17.280
(Sec. 4)	Age	15.230	0.580
	Socio-ec. Index	-	-
	Perc. High	0.337	0.473
	Perc. Med.	0.403	0.491
	Gender (Female=1)	0.533	0.499
	Foreign	0.127	0.333
	Secondary 5	0.044	0.205
	Number of observations	55057	
	Number of groups	382	
	Avg. size of true groups	144.6	134.8
	Avg. size of observed groups	144.1	134.5

† The groups include only students for whom *Foreign*=0, that is, whose language of instruction is the same as the mother tongue and the language spoken at home.
†† Math refers to Math 514 (Secondary 5 regular course).

Table 2: Peer Effects on Scholar Achievement[a]
Naive OLS

	French	French2[b]	Sciences	Math	History
Endogenous effect	0.845**	0.828**	0.899**	0.831**	0.858**
	(0.015)	(0.016)	(0.011)	(0.018)	(0.012)
Individual effects					
Constant	102.916**	127.535**	115.048**	96.252**	137.010**
	(10.114)	(10.994)	(9.450)	(5.588)	(7.456)
Age	-7.440**	-8.112**	-7.819**	-4.376**	-7.294**
	(0.152)	(0.170)	(0.150)	(0.214)	(0.140)
Socio-ec. (High)	1.344**	1.350**	1.592**	1.562**	2.126**
	(0.209)	(0.224)	(0.228)	(0.397)	(0.227)
Socio-ec. Index (Medium)	0.716**	0.763**	0.676**	0.833**	0.893**
	(0.182)	(0.195)	(0.198)	(0.329)	(0.197)
Gender (Female=1)	3.993**	4.182**	0.252*	0.750**	-1.512**
	(0.127)	(0.132)	(0.138)	(0.233)	(0.136)
Foreign	-2.368**	-	2.150**	0.047	0.567**
	(0.263)	(-)	(0.255)	(0.460)	(0.254)
Secondary 5	9.408**	8.967**	1.294**	6.798**	2.557**
	(0.736)	(0.795)	(0.310)	(0.659)	(0.348)
Contextual effects					
Age	2.101**	1.391**	0.781	-0.987**	-1.038**
	(0.581)	(0.634)	(0.632)	(0.305)	(0.471)
Socio-ec. (High)	-1.841**	-1.791**	-2.239**	-2.003**	-2.896**
	(0.351)	(0.367)	(0.385)	(0.602)	(0.377)
Socio-ec. Index (Medium)	-1.227**	-1.219**	-0.902**	-0.320	-1.179**
	(0.384)	(0.394)	(0.416)	(0.653)	(0.414)
Gender (Female=1)	-4.324**	-4.292**	-1.435**	-0.028	0.814
	(0.528)	(0.564)	(0.497)	(1.089)	(0.534)
Foreign	3.416**	-	-0.667	2.247**	0.913**
	(0.435)	(-)	(0.455)	(0.807)	(0.445)
Secondary 5	-14.306**	-15.121**	9.992**	-5.706**	10.037**
	(1.469)	(1.734)	(1.567)	(1.396)	(1.897)

Notes:
Standard errors in parentheses
** indicates 5% significance level
* indicates 10% significance level
[a]The dependent variable is the score on June 2005 provincial secondary exams.
[b]The groups include only students for whom *Foreign*=0, that is, whose language of instruction is the same as the mother tongue and the language spoken at home.

27

Table 3: Peer Effects on Student Achievement[a]
Conditional Maximum Likelihood

	French	French2[b]	Science	Math	History
Endogenous effect	0.330	0.511*	-0.239	0.827**	0.650**
	(0.328)	(0.312)	(0.234)	(0.249)	(0.272)
Individual effects					
Age	-7.999**	-8.707**	-8.293**	-4.868**	-7.942**
	(0.162)	(0.174)	(0.151)	(0.271)	(0.151)
Socio-ec. Index (High)	1.423**	1.505**	1.609**	2.112**	2.019**
	(0.245)	(0.267)	(0.268)	(0.500)	(0.261)
Socio-ec. Index (Medium)	0.669**	0.772**	0.785**	1.189**	0.795**
	(0.220)	(0.241)	(0.230)	(0.435)	(0.234)
Gender (Female=1)	3.807**	4.114**	0.319**	1.018**	-1.641**
	(0.162)	(0.160)	(0.158)	(0.301)	(0.159)
Foreign	-2.596**	-	2.095**	-0.081	0.807**
	(0.279)	(-)	(0.278)	(0.548)	(0.284)
Secondary 5	10.520**	9.458**	1.652**	6.474**	3.127**
	(0.676)	(0.673)	(0.328)	(0.767)	(0.399)
Contextual effects					
Age	-39.302**	-34.174**	-19.554**	0.838	-31.569**
	(10.991)	(8.892)	(8.338)	(7.382)	(9.471)
Socio-ec. Index (High)	16.506	23.069	8.934	29.310*	-6.411
	(17.534)	(18.110)	(22.453)	(15.580)	(18.948)
Socio-ec. Index (Medium)	-4.916	2.251	22.144	18.246	-6.750
	(16.875)	(17.433)	(17.782)	(13.726)	(18.566)
Gender (Female=1)	-25.019*	-8.069	14.855	15.558	-11.821
	(14.397)	(11.480)	(12.178)	(9.491)	(12.413)
Foreign	-26.681*	-	-8.851	-2.654	29.208
	(15.865)	(-)	(16.952)	(12.143)	(18.008)
Secondary 5	167.966**	67.021**	-0.343	-6.080	24.068
	(41.189)	(28.798)	-19.955	(26.056)	(21.167)
Log-likelihood	-162548.55	-143890.84	-226078.18	-62420.96	-226216.11

Notes:

Standard errors in parentheses

** indicates 5% significance level

* indicates 10% significance level

[a]The dependent variable is the score on June 2005 provincial secondary exams.

[b]The groups include only students for whom *Foreign*=0, that is, whose language of instruction is the same as the mother tongue and the language spoken at home.

Table 4: Peer Effects on Scholar Achievement[a]

OLS Estimation with Group Fixed Effects and no endogenous effect[b]

	French	French2[c]	Sciences	Math	History
Individual effects					
Age	-7.572**	-8.169**	-7.994**	-4.565**	-7.451**
	(0.150)	(0.168)	(0.146)	(0.215)	(0.139)
Socio-ec. (High)	1.417**	1.635**	1.756**	1.933**	2.001**
	(0.233)	(0.267)	(0.257)	(0.442)	(0.245)
Socio-ec. Index (Medium)	0.660**	0.856**	0.832**	1.217**	0.802**
	(0.208)	(0.243)	(0.231)	(0.378)	(0.222)
Gender (Female=1)	3.831**	4.084**	0.328**	0.975**	-1.540**
	(0.171)	(0.160)	(0.162)	(0.270)	(0.159)
Foreign	-2.555**	-	2.206**	0.104	0.977**
	(0.284)	(-)	(0.280)	(0.507)	(0.275)
Secondary 5	10.087**	9.469**	1.513**	6.648**	2.900**
	(0.789)	(0.817)	(0.327)	(0.765)	(0.370)
Contextual effects					
Age	6.615**	9.093**	12.133**	7.036**	14.267**
	(2.286)	(1.642)	(1.106)	(1.489)	(0.959)
Socio-ec. (High)	15.215	38.375**	30.030	19.537**	-13.348
	(13.792)	(17.336)	(18.261)	(9.417)	(12.885)
Socio-ec. Index (Medium)	-6.389	11.013	26.551	20.022**	-9.021
	(12.893)	(16.853)	(17.505)	(8.462)	(14.325)
Gender (Female=1)	-20.917	-11.185	14.660	13.342**	0.238
	(16.487)	(11.264)	(13.138)	(6.220)	(12.001)
Foreign	-23.074*	-	5.688	3.540	49.751**
	(12.190)	(-)	(14.438)	(7.625)	(12.064)
Secondary 5	106.712**	59.504**	-2.889	4.864	13.204
	(36.935)	(24.882)	(18.320)	(23.402)	(17.280)

Notes:

Standard errors in parentheses

** indicates 5% significance level

* indicates 10% significance level

[a]The dependent variable is the score on June 2005 provincial secondary exams.

[b]The groups include only students for whom *Foreign*=0, that is, whose language of instruction is the same as the mother tongue and the language spoken at home.

Table 5: Peer Effects on Student Achievement[a]

Likelihood Ratio Tests - Contextual Effects (Estimates from the restricted model)

	French	French2[b]	Science	Math	History
Endogenous effect	0.438	0.644**	-0.070	0.779**	0.701**
	(0.326)	(0.310)	(0.229)	(0.239)	(0.269)
Individual effects					
age	-7.698**	-8.385**	-8.150**	-4.882**	-7.706**
	(0.136)	(0.151)	(0.137)	(0.220)	(0.134)
Socio-ec. Med	1.324**	1.332**	1.539**	1.534**	2.072**
	(0.209)	(0.222)	(0.227)	(0.401)	(0.228)
Socio-ec. High	0.702**	0.750**	0.637**	0.794**	0.848**
	(0.179)	(0.191)	(0.195)	(0.334)	(0.196)
Gender	3.991**	4.189**	0.224	0.708**	-1.562**
	(0.126)	(0.132)	(0.137)	(0.238)	(0.137)
Foreign	-2.372**	-	2.154**	-0.001	0.575**
	(0.244)	(-)	(0.246)	(0.459)	(0.246)
Secondary 5	9.222**	8.875**	1.616**	6.597**	2.929**
	(0.583)	(0.608)	(0.302)	(0.637)	(0.366)
Non-restricted log-likelihood	-162548.55	-143890.84	-226078.18	-62420.961	-226216.11
Restricted log-likelihood	-162567.89	-143905.42	-226082.99	-62424.425	-226223.58
Likelihood ratio	38.691259	29.162129	9.6197715	6.9264363	14.939641

Notes:

Standard errors in parentheses

** indicates 5% significance level

* indicates 10% significance level

[a]The dependent variable is the score on June 2005 provincial secondary exams.

[b]The groups include only students for whom *Foreign*=0, that is, whose language of instruction is the same as the mother tongue and the language spoken at home.

Table 6: Peer Effects on Student Achievement[a]

2SLS Estimation with Group Fixed Effect[b]

	French	French2[c]	Sciences	Math	History
Endogenous effect	1.378	1.580	-0.509	-0.037	0.787
	(1.468)	(1.456)	(0.764)	(0.477)	(0.980)
Individual effects					
Age	-7.690**	-8.359**	-7.962**	-4.606**	-7.609**
	(0.197)	(0.232)	(0.167)	(0.228)	(0.163)
Socio-ec. Index (High)	1.373**	1.458**	1.754**	1.836**	2.041**
	(0.242)	(0.248)	(0.250)	(0.423)	(0.248)
Socio-ec. Index (Medium)	0.661**	0.716**	0.826**	1.069**	0.803**
	(0.221)	(0.234)	(0.219)	(0.365)	(0.221)
Gender (Female=1)	3.871**	4.103**	0.333**	0.965**	-1.553**
	(0.164)	(0.159)	(0.159)	(0.265)	(0.157)
Foreign	-2.514**	-	2.128**	-0.005	0.716**
	(0.282)	(-)	(0.270)	(0.496)	(0.276)
Secondary 5	9.516**	9.022**	1.415**	6.674**	2.910**
	(0.781)	(0.840)	(0.327)	(0.741)	(0.390)
Contextual effects					
Age	4.205	4.669	13.496**	6.713**	8.552**
	(4.845)	(4.268)	(3.050)	(1.712)	(4.036)
Socio-ec. Index (High)	7.364	17.348	30.997*	15.962**	-6.246
	(17.305)	(13.585)	(16.678)	(7.641)	(15.620)
Socio-ec. Index (Medium)	-7.103	-5.423	26.344*	13.501*	-8.047
	(16.813)	(16.138)	(13.908)	(7.555)	(14.598)
Gender (Female=1)	-21.310*	-15.413	15.637	13.237**	0.567
	(12.261)	(10.472)	(12.202)	(5.808)	(11.708)
Foreign	-15.732	-	-2.232	-0.065	19.385
	(12.571)	(-)	(11.449)	(7.189)	(12.903)
Secondary 5	40.184	16.768	-17.370	7.825	2.537
	(36.380)	(37.192)	(14.470)	(21.360)	(23.060)
Sargan Test	23.52	14.17	0.54	1.40	5.35
[*p-value*]	[0.00]	[0.03]	[1.00]	[0.97]	[0.50]
Stock and Yogo Test	706.84	612.93	1055.92	464.43	660.40
[*Critical Value for b=0.05*					
at sign. level of 5%]	[18.37]	[18.37]	[18.37]	[18.37]	[18.37]

Notes:

Standard errors in parentheses

** indicates 5% significance level

* indicates 10% significance level

[a]The dependent variable is the score on June 2005 provincial secondary exams.

[b]The groups include only students for whom *Foreign*=0, that is, whose language of instruction is the same as the mother tongue and the language spoken at home.

Table 7: Peer Effects on Student Achievement[a]

Generalized 2SLS Estimation[b]

	French	French2[c]	Sciences	Math	History
Endogenous effect	-2.104	-3.882	-0.015	-0.162	-2.753
	(3.619)	(2.661)	(0.734)	(0.465)	(1.717)
Individual effects					
Age	-7.390**	-7.783**	-8.012**	-4.582**	-7.306**
	(0.348)	(0.322)	(0.165)	(0.227)	(0.203)
Socio-ec. Index (High)	1.542**	1.540**	1.718**	1.844**	2.222**
	(0.293)	(0.237)	(0.251)	(0.421)	(0.250)
Socio-ec. Index (Medium)	0.867**	0.914**	0.803**	1.080**	0.921**
	(0.293)	(0.232)	(0.220)	(0.362)	(0.219)
Gender (Female=1)	3.770**	4.046**	0.318**	0.966**	-1.536**
	(0.186)	(0.153)	(0.161)	(0.264)	(0.150)
Foreign	-2.568**	-	2.144**	0.006	0.642**
	(0.283)	(-)	(0.273)	(0.494)	(0.268)
Secondary 5	9.797**	9.663**	1.471**	6.701**	2.560**
	(0.799)	(0.837)	(0.327)	(0.739)	(0.393)
Contextual effects					
Age	15.211	21.620**	11.514**	7.014**	22.639**
	(11.684)	(8.190)	(2.967)	(1.646)	(6.808)
Socio-ec. Index (High)	31.802	29.153**	25.140	16.299**	25.748
	(30.455)	(11.765)	(16.904)	(7.431)	(18.608)
Socio-ec. Index (Medium)	21.574	21.694	23.091	14.010*	10.015
	(31.981)	(17.288)	(14.514)	(7.338)	(14.788)
Gender (Female=1)	-20.267*	0.403	13.639	13.265**	-1.936
	(11.040)	(12.140)	(12.713)	(5.676)	(9.533)
Foreign	-28.226	-	-1.320	0.354	9.226
	(18.593)	(-)	(12.377)	(7.041)	(12.411)
Secondary 5	79.885*	90.818**	-12.062	9.953	-31.537
	(46.845)	(42.242)	(15.235)	(21.115)	(20.155)

Notes:

Standard errors in parentheses

** indicates 5% significance level

* indicates 10% significance level

[a]The dependent variable is the score on June 2005 provincial secondary exams.

[b]The groups include only students for whom *Foreign*=0, that is, whose language of instruction is the same as the mother tongue and the language spoken at home.

Table 8: Group Size Variation

Simulations using CML and G2SLS

Avg. Group size	Group sizes Range	Endogenous effects CML Avg. Coeff	SE	G2SLS Avg. Coeff	SE	Contextual effects - Age CML Avg. Coeff	SE	G2SLS Avg. Coeff	SE	Contextual effects - Gender CML Avg. Coeff	SE	G2SLS Avg. Coeff	SE
10	[3;17]	0.35	0.00	0.35	0.02	-40.01	0.25	-40.00	0.25	-25.01	0.33	-25.00	0.32
10	[5;15]	0.35	0.00	0.35	0.10	-40.00	0.35	-39.99	0.44	-24.99	0.74	-24.96	1.03
10	[7;13]	0.35	0.02	0.37	0.42	-40.00	0.53	-40.04	0.98	-25.01	1.50	-25.18	3.94
10	[9;11]	0.57	0.38	-10.39	5.74E+07	-40.27	1.79	-31.98	9.31E+07	-26.97	5.78	47.59	5.20E+08
20	[3;37]	0.35	0.00	0.35	0.03	-40.01	0.27	-40.00	0.26	-25.03	0.44	-25.01	0.42
20	[8;32]	0.35	0.02	0.35	0.29	-40.00	0.50	-40.01	0.67	-25.02	1.10	-24.99	2.13
20	[13;27]	0.35	0.09	0.48	1.43	-39.95	1.23	-39.55	4.51	-25.04	2.11	-25.85	9.37
20	[18;22]	0.94	1.56	-23.21	4.90E+03	-37.98	5.37	-120.51	1.63E+04	-28.55	8.47	122.01	3.09E+04
40	[3;77]	0.35	0.00	0.35	0.05	-39.98	0.41	-39.98	0.38	-25.03	0.65	-25.02	0.59
40	[12;68]	0.36	0.07	0.38	0.74	-39.97	1.42	-39.86	3.52	-25.05	1.66	-25.17	4.29
40	[21;59]	0.39	0.20	0.72	3.15	-39.85	2.76	-38.01	17.17	-25.14	2.67	-26.84	16.43
40	[30;50]	0.72	0.85	-64.96	7.03E+04	-37.92	5.82	-413.66	4.02E+05	-26.93	5.30	309.80	3.59E+05
40	[39;41]	1.00	155.98	-42.37	4.25E+07	-36.25	78.67	-305.31	2.32E+08	-26.28	69.22	190.47	1.93E+08
80	[3-157]	0.35	0.01	0.35	0.07	-39.99	0.68	-39.99	0.62	-25.05	0.98	-25.04	0.91
80	[18-142]	0.43	0.19	0.45	1.84	-39.55	2.93	-39.46	11.18	-25.40	2.49	-25.44	9.09
80	[33-127]	0.57	0.46	1.51	6.71	-38.54	4.87	-32.25	44.27	-25.96	3.66	-30.25	30.77
80	[48-112]	0.87	1.20	4.78	281.37	-36.47	8.02	-9.84	1.92E+03	-27.10	5.75	-44.93	1.23E+04
80	[63-97]	1.00	5.27	-17.05	1.26E+04	-35.75	17.05	-159.42	8.76E+04	-27.74	11.97	52.89	5.53E+04
120	[3-237]	0.36	0.01	0.36	0.08	-39.99	0.99	-39.98	0.89	-25.10	1.50	-25.12	1.38
120	[28-212]	0.64	0.51	0.74	4.94	-38.14	5.22	-37.38	33.42	-26.34	3.79	-26.75	22.26
120	[53-187]	0.89	1.30	6.66	23.63	-35.99	8.69	5.19	168.02	-27.25	5.85	-51.90	101.50
120	[78-162]	1.00	3.85	-10.58	3.98E+03	-35.34	15.16	-116.30	2.88E+04	-27.65	9.94	21.54	1.69E+04
120	[103-137]	1.00	25.15	-35.80	1.40E+06	-35.32	39.00	-306.98	1.01E+07	-28.20	25.29	140.78	5.64E+06

Table 9: Simulations Calibrated on French Sample

(1000 replications)

	CML	2SLS	G2SLS	OLS
Endogenous effect	0.391	-0.873	0.495	-33.571
	(0.101)	(0.852)	(167.702)	(3.688)
Individual effects				
Age	-8.002	-7.920	-8.006	-5.758
	(0.145)	(0.149)	(10.021)	(0.545)
Gender (Female=1)	3.798	3.822	3.828	4.480
	(0.147)	(0.139)	(1.693)	(0.554)
Exogenous effects				
Age	-39.996	-38.085	-39.540	17.373
	(9.996)	(7.579)	(167.394)	(76.788)
Gender (Female=1)	-25.329	-16.703	-21.857	210.526
	(10.733)	(10.092)	(692.625)	(74.714)

Notes: Average standard errors are in parentheses. The group sizes are calibrated on our French sample. $\sigma^2 = \hat{\sigma}^2$ (calibrated)$= 154.704$. The other parameters are calibrated as in Table 8.