

DO PEERS AFFECT STUDENT ACHIEVEMENT? EVIDENCE FROM CANADA USING GROUP SIZE VARIATION

VINCENT BOUCHER,^a YANN BRAMOULLÉ,^b HABIBA DJEBBARI^c AND BERNARD FORTIN^{d*}

^a CIREQ and Department of Economics, Université de Montréal, Montréal, Québec, Canada

^b CIRPÉE and Department of Economics, Université Laval, Québec, Canada; AMSE, Aix-Marseille, France

^c CIRPÉE and Department of Economics, Université Laval, Québec, Canada; AMSE, Aix-Marseille, France; IZA

^d CIRPÉE and Department of Economics, Université Laval, Québec, Canada; CIRANO and IZA

SUMMARY

We provide the first empirical application of a new approach proposed by Lee (*Journal of Econometrics* 2007; **140** (2), 333–374) to estimate peer effects in a linear-in-means model when individuals interact in groups. Assuming-sufficient group size variation, this approach allows to control for correlated effects at the group level and to solve the simultaneity (reflection) problem. We clarify the intuition behind identification of peer effects in the model. We investigate peer effects in student achievement in French, Science, Mathematics and History in secondary schools in the Province of Québec (Canada). We estimate the model using conditional maximum likelihood and instrumental variables methods. We find some evidence of peer effects. The endogenous peer effect is large and significant in Mathematics but imprecisely estimated in the other subjects. Some contextual peer effects are also significant. In particular, for most subjects, the average age of peers has a negative effect on own test score. Using calibrated Monte Carlo simulations, we find that high dispersion in group sizes helps with potential issues of weak identification. Copyright © 2012 John Wiley & Sons, Ltd.

Received 26 January 2010; Revised 10 June 2012



Supporting information may be found in the online version of this article.

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 tracking, racial integration and school vouchers (for a recent survey, see Epple and Romano, 2011). However, despite a growing literature on the subject, the evidence regarding the magnitude of peer effects on student achievement is mixed (e.g. Sacerdote, 2001; Hanushek *et al.*, 2003; Stinebrickner and Stinebrickner, 2006; Ammermueller and Pischke, 2009). This lack of consensus partly reflects various econometric issues that any empirical study on peer effects must address. Identifying and estimating peer effects raises three basic challenges. First, the relevant peer groups must be determined. Who interacts with whom? Second, peer effects must be identified from confounding factors. In particular, 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. Simultaneity, also called the *reflection problem* by Manski (1993), may prevent separating contextual effects, i.e. the influence of peers' characteristics, from the endogenous effect, i.e. the influence of peers' outcome. This issue is important since only the endogenous effect is the source of a *social multiplier*. Researchers have adopted 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.

* Correspondence to: Bernard Fortin, Department of Economics, Université Laval, J.A. De Sève Building, 1025, avenue des Sciences-Humaines, Québec, Canada, G1V 0A6. E-mail: bernard.fortin@ecn.ulaval.ca

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. In principle, the approach is promising, as it allows to solve the problem of correlated effects and the reflection problem with standard observational (non-experimental) data. Moreover, the exclusion restrictions imposed by the model are explicitly derived from its structural specification and provide natural instruments. The econometric model does rely on a number of crucial assumptions, however, which makes its confrontation to real data particularly important. We empirically assess the approach using original administrative data on test scores at the end of secondary school in the Canadian Province of Québec. We investigate the presence of peer effects in student achievement in Mathematics, Science, French, and History. In the process, we also provide new economic insights regarding the sources of identification in the model. This matters in particular in assessing its robustness to alternative (nonlinear) approaches.

The econometric model relies on three key assumptions. First, individuals interact in groups known to the modeler. This means that the population of students is partitioned into groups (e.g. classes, grade levels) and that students are affected by all their peers in their groups but by none outside of it. This assumption is typical in studies of academic achievement but clearly arises from data constraints. Second, each individual's peer group is everyone in his group *excluding* himself. While this assumption seems innocuous and has been used in most empirical studies, it is a key source of identification in the model, as will become clear below. In fact, it is a main source of difference between Manski's (1993) and Lee's models. Manski's approach can be interpreted as one in which each individual's peer group *includes* himself.¹ Third, individual outcome is determined by a linear-in-means model with group fixed effects. Thus the test score of a student is affected 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 identified in such a framework when there are sufficient groups of different sizes. One important contribution of our paper is to clarify the economic intuition behind identification. Regarding the estimation of parameters, one potentially important limitation of the method, however, is that convergence in distribution of the peer effect estimates may occur at low rates when the average group size is large relative to the number of groups in the sample (Lee, 2007). This is also intuitive: excluding the individual or not from his peer group does not change much when its size is relatively large.

Here two remarks are in order. First, these results are to be distinguished from the idea that the group size is a factor in a school's production function (e.g. Krueger, 2003). In Lee's model, the effects of group sizes which are separable from the peer effects are controlled for by fixed effects in the structural model. Second, Lee's identification method differs from the *variance contrast* approach developed by Graham (2008). The basic idea in this approach is that peer effects will induce intra-group dependencies in behavior that introduce variance restrictions on the error terms. These restrictions are used to identify the composite (endogenous + contextual) social interaction effects under the assumption that the variance matrix parameters are independent of the reference group size.

We use administrative data on academic achievement for a large sample of secondary schools in the Province of Québec obtained from the Ministry of Education, Recreation and Sports (MERS). Our dependent variables are individual scores on four standardized tests taken in June 2005 (Mathematics, Science, French and History) by fourth- and fifth-grade secondary school students. All fourth- and fifth-grade students in the province must pass these tests to graduate. One advantage of these data is that all candidates in the province take the same exams, no matter what their school and location. This feature effectively allows us to consider test scores as draws from 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 Québec schools which, over the four subjects, yields 194,553 test scores for

¹ More precisely, Manski studies a social interactions model which, in terms of identification, has the same properties as a model where individuals interact in groups and each individual is included in his peer group (see Bramoullé *et al.*, 2009).

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 modeling 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 results. Following Lee (2007), we estimate the model in two ways: through generalized instrumental variables (IV) and, under stronger parametric conditions, through conditional maximum likelihood robust to non-normal disturbances (pseudo CML).

Our results are mixed though consistent with the model. We do provide evidence of some endogenous and contextual peer effects. Based on pseudo CML estimates, we find that the endogenous peer effect is positive, significant and quite high in Mathematics (0.83). Moreover it is within the range of previous estimates (see Sacerdote, 2011, for a recent survey). However, the effect is smaller and non-significant in History (0.64), French (0.30) and Science (-0.23).² Endogenous peer effects estimates obtained from IV methods are highly imprecise with our data, even in Mathematics. The higher precision of our pseudo CML estimates is consistent with results in Lee (2007) showing that CML estimators are asymptotically more efficient than IV estimators. As regards contextual peer effects, we find evidence that some of them matter, based on both pseudo CML and IV estimators. For instance, results from pseudo CML indicate that interacting with older students (a proxy for repeaters) has a negative effect on own test score in all subjects except Mathematics (not significant).

It is remarkable that even with large average group size relative to the number of groups we are able to identify some peer effects. However, there is also much dispersion in group sizes within our samples. We suspect that this helps identification. We study this issue systematically through Monte Carlo simulations. We find that indeed increasing group size dispersion has a positive impact on the precision of estimates.

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.

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.³

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, for example, Calvó-Armengol *et al.*, 2009; Lin, 2010).⁴ For this reason, many studies focus on the grade-within-school level (e.g. Hanushek *et al.*, 2003; Angrist and Lang, 2004). Other studies analyze peer effects at the classroom level (e.g. Kang, 2007; Ammermueller and Pischke, 2009). 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

² The effects 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.

³ For two recent comprehensive surveys on peer effects in education, see Sacerdote (2011) and Epple and Romano (2011).

⁴ Bramoullé *et al.* (2009) determine conditions under which endogenous and contextual peer effects are identified when students interact through a social network known by the modeler and when correlated effects are fixed within subnetworks. See also Section 3.4.2. in this paper.

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.⁵

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. Sacerdote, 2001; Zimmerman, 2003; Kang, 2007). Results on the impact of contextual effects using randomly assigned roommates as peers are usually low though significant. However, Stinebrickner and Stinebrickner (2006) have argued that these studies tend to underestimate true peer effects as the true influence of roommates is unclear. A second strategy uses observational data to estimate peer effects. This approach is usually based on two assumptions. First, fixed effects allow correlated effects to be taken into account. With cross-section data, these effects are usually defined at a level higher than peer groups. Otherwise, peer effects are absorbed in these effects and cannot therefore be identified. For instance, Ammermueller and Pischke (2009) introduce *school* fixed effects to estimate peer effects at the *class* level for fourth graders in six European countries. Contrary to this approach, our model allows inclusion of fixed effects at the *peer group* level even with cross-section data. This is so because each student within a group has his own reference group (since he is excluded from it). The second assumption is that one observes exogenous shocks to peer group composition which allow identification of a composite (endogenous + contextual) peer effect. The strategy uses either cross-section or panel data. With cross-section data, demographic variations across grades but within schools are usually exploited (see Bifulco *et al.*, 2011). With panel data, demographic variations across cohorts but within school grades are usually exploited (see Hanushek *et al.*, 2003).

The reflection problem is handled using two main strategies. In most papers, no solution for this difficult problem is provided. Rather, 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). Note, however, that a number of these papers (often implicitly) assume that there are no contextual effects. In this case, the composite parameter(s) allow(s) to identify the endogenous peer effect. 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). The problem here is to choose suitable instruments. 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 naturally derived from the structure of the model.

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 both endogenous and contextual peer effects may be fully identified even with observational data in cross-section.

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 $\{i = 1, \dots, M\}$ is supposed to be partitioned into groups of peers indexed by $r = 1, \dots, R$. Let M_r be the r th 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

⁵ In fact, at the end of secondary level, classes and teachers are usually different, depending on the subject matter taught.

group r is excluded from his own reference group. Let M_{ri} be student i 's group of peers, of size $m_r - 1$. 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 expository purposes, the model is first presented with a unique characteristic ($K = 1$), defined by his family socioeconomic background. 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 .

We do not change any other assumption of the linear-in-means model. In particular, we assume that a student'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 basic structural equation is given by

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

where β captures the endogenous effect, γ the individual effect and δ is the contextual effect. Observe that equation (1) can be derived from the first-order conditions of a choice-theoretic non-cooperative (Nash) model where each student's performance is obtained from the maximization of his quadratic utility function, which depends on his individual characteristics, his performance and his reference group's mean performance and mean characteristics.

Importantly, we assume strict exogeneity of m_r and $\{x_{ri} : i = 1, \dots, m_r\}$ conditional on the unobserved effect α_r , i.e. $\mathbb{E}(\varepsilon_{ri} | \mathbf{X}_r, m_r, \alpha_r) = 0$. This exogeneity assumption can notably accommodate 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, α_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: $\mathbb{E}(\varepsilon_{ri} | \mathbf{X}_r, m_r, \alpha_r) = 0$. We maintain this assumption throughout our analysis.

To eliminate group-invariant correlated effects, we next apply a *within* transformation to equation (1). In particular, as we noted above, when the effect of group size is separable from peer and individual effects, it is captured by α_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. 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. The within reduced-form equation for students in the r th group can be written as

$$y_{ri} - \bar{y}_r = \frac{\gamma - \frac{\delta}{m_r - 1}}{1 + \frac{\beta}{m_r - 1}} (x_{ri} - \bar{x}_r) + \frac{1}{1 + \frac{\beta}{m_r - 1}} (\varepsilon_{ri} - \bar{\varepsilon}_r) \quad (2)$$

where means \bar{y}_r , \bar{x}_r and $\bar{\varepsilon}_r$ are computed over *all* students in the group. Now assume that $\gamma\beta + \delta \neq 0$. 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 .⁶

⁶ It is easy to show that when $\gamma\beta + \delta = 0$, only γ is identified.

3.2. Interpretation of Identification

The fact that the parameters of the structural within equation (2) may be fully identified is quite surprising, and deserves some elaboration. Indeed, under the alternative assumption that means are inclusive, i.e. $i \in M_{ri}$, peers are the same for everyone in a group $M_{ri} = M_r$, and peer effects cannot be separated out from group fixed effects. Thus somehow assuming that the individual is excluded from his own peer group allows us to solve two difficult identification problems: distinguishing true peer effects from correlated effects and further distinguishing endogenous from contextual peer effects. Intuitively, where does identification come from?

Suppose first that the endogenous effect is absent $\beta = 0$. Note that each individual has different peers: $i \neq k$ implies that $M_{ri} \neq M_{rk}$. A first key observation is that, within a group, individual attributes x_i are perfectly negatively correlated with mean peer attributes $\left(\sum_{j \in M_{ri}} x_j\right)/(m_r - 1)$.⁷ Thus students with an ability above average necessarily have peers with a mean ability below average, and vice versa. If the individual and the contextual effects γ and δ are positive, this negative correlation tends to reduce the dispersion in outcomes. In such a group setting, peer effects lower the difference in achievement between high- and low-ability students.⁸ Formally, the impact of the difference in attributes on the difference in outcomes changes from γ to $\gamma - \delta/(m_r - 1)$ when introducing peer effects (see, for example, equation (2)). So variations in group sizes can be used to identify contextual peer effects. The second key observation is that this reduction is stronger in smaller groups. The variance in mean peer attributes is simply higher in smaller groups, reflecting the relatively larger effect of excluding one individual from the mean. And as group size increases, mean peer attributes converge to the group mean, and peer effects have increasingly less bite on how differences in covariates affect differences in outcomes.

Next, consider the reflection problem. Observe that outcomes are subject to a similar negative correlation: within a group, students with grades above average necessarily have peers with grades below average. Thus, if $\beta > 0$, endogenous peer effects lead to a further reduction in outcome dispersion. However, simultaneity now implies that this decrease in impact is *nonlinear* in the peer coefficient: from $\gamma - \delta/(m_r - 1)$ to $(\gamma - \delta/(m_r - 1))/(1 + \beta/(m_r - 1))$ (see equation (2)). The difference in the shapes of impact reduction can then be used to identify endogenous from contextual peer effects.

Finally, this understanding is useful in assessing the robustness of the identification strategy to changes in the econometric model. In particular, it is easy to see that if $x_i < x_k$ then the distribution of attributes in i 's peer group M_{ri} first-order stochastically dominates the distribution in M_{rk} . Thus identification is likely to hold, in general, if we replace the mean in equation (1) by the median, the variance, or many other moments of the distribution.⁹

3.3. Treatment of Missing Values

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 use a correction first developed by Davezies *et al.* (2009) 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 $n_r \leq m_r$ and $\sum_{r=1}^R n_r = N$. We assume that a student's decision to postpone exam-taking is random or depends on the observable strictly

⁷ To see this, observe that $\sum_{j \in M_{ri}} x_j = \left(\sum_{j \in M_r} x_j\right) - x_i$. So if $x_i < x_k$ then $\frac{1}{m_r - 1} \sum_{j \in M_{ri}} x_j > \frac{1}{m_r - 1} \sum_{j \in M_{rk}} x_j$.

⁸ In contrast, if $\gamma > 0$ and $\delta < 0$ *amplify* the dispersion in outcomes.

⁹ Of course, one has to address a basic modeling question first, i.e. whether the implied model is coherent. A model has this property when a specific nonlinear structure generates a unique solution for outcomes.

exogenous variables, conditional on the fixed group effect. We show how to adapt Lee's analysis to this more general setting. Let L_r be the complement of N_r , i.e. $L_r = M_r - N_r$.¹⁰ The structural equation becomes

$$y_{ri} = \tilde{\alpha}_r + \beta \frac{\sum_{j \in N_r} y_{rj}}{m_r - 1} + \gamma x_{ri} + \delta \frac{\sum_{j \in N_r} x_{rj}}{m_r - 1} + \varepsilon_{ri}, \quad \mathbb{E}(\varepsilon_{ri} | \mathbf{X}_r, m_r, \alpha_r) = 0 \quad (3)$$

where i now denotes an observed individual in the sample (but not any one in the r th group) and $\tilde{\alpha}_r = \alpha_r + \beta \frac{\sum_{j \in L_r} y_{rj}}{m_r - 1} + \delta \frac{\sum_{j \in L_r} x_{rj}}{m_r - 1}$ is the new group fixed effect. Under our assumptions, estimators are consistent, even if we do not observe test scores for all students in each group. Moreover, effects stemming from unobserved individuals are the same for all the individuals observed in the sample from the r th group. They are therefore picked up by the group fixed effect. Using the within transformation, one obtains the same equation as (2) but where means \bar{y}_r , \bar{x}_r and $\bar{\varepsilon}_r$ are computed only over all *observed* students in the group.

3.4. Estimation Methods

3.4.1. CML Estimator

We consider estimation under both pseudo conditional maximum likelihood (or CML) and instrumental variables (or IV) identification conditions.

To present pseudo CML and IV estimators, it is easier to express equation (3) in matrix notations. We now allow for any number of characteristics, so that γ is a $K \times 1$ vector of individual effects and δ 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 modeled as an $N \times N$ block-diagonal matrix $\mathbf{G} = \text{Diag}(\mathbf{G}_1, \dots, \mathbf{G}_R)$, 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, $\mathbf{G}_r = \frac{1}{m_r - 1} (\mathbf{I}_{n_r} \mathbf{I}_{n_r}' - \mathbf{I}_{n_r})$, where \mathbf{I}_{n_r} is a $n_r \times 1$ vector of ones and \mathbf{I}_{n_r} is the identity matrix of dimension n_r . Equation (3) can be rewritten in matrix form as follows:

$$\mathbf{y}_r = \mathbf{I}_{n_r} \tilde{\alpha}_r + \beta \mathbf{G}_r \mathbf{y}_r + \mathbf{X}_r \gamma + \mathbf{G}_r \mathbf{X}_r \delta + \varepsilon_r \quad (4)$$

where $\mathbb{E}(\varepsilon_r | \mathbf{X}_r, \mathbf{G}_r, \tilde{\alpha}_r) = 0$.

Applying the operator matrix $\mathbf{J}_r = \mathbf{I}_{n_r} - \frac{1}{n_r} \mathbf{I}_{n_r} \mathbf{I}_{n_r}'$ allows us to obtain deviations with respect to the mean for the observed group members. Pre-multiplying equation (4) by \mathbf{J}_r eliminates the group fixed effect and yields

$$\mathbf{J}_r \mathbf{y}_r = \beta \mathbf{J}_r \mathbf{G}_r \mathbf{y}_r + \mathbf{J}_r \mathbf{X}_r \gamma + \mathbf{J}_r \mathbf{G}_r \mathbf{X}_r \delta + \mathbf{J}_r \varepsilon_r \quad (5)$$

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

$$\frac{m_r - 1 + \beta}{m_r - 1} \mathbf{y}_r^* = \mathbf{X}_r^* \frac{(m_r - 1)\gamma - \delta}{m_r - 1} + \varepsilon_r^*$$

which is equivalent to equation (2).

To derive the pseudo CML estimator, we assume (possibly wrongly) 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 , the pseudo density of \mathbf{y}_r^* is a multivariate normal distribution with

¹⁰ If N_{ri} denotes the group of peers of student i , we also have $L_r = M_{ri} - N_{ri}$.

mean $\mathbf{X}_r^* \frac{(m_r-1)\gamma-\delta}{m_r-1+\beta}$ and variance $\left(\sigma \frac{m_r-1}{m_r-1+\beta}\right)^2 \mathbf{J}_r$.¹¹ The pseudo log-likelihood function to be maximized can then be expressed as follows:

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

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 $r = 1, \dots, R$. Under the assumption that the ε_{ir} s are correctly specified and i.i.d. $\mathbf{N}(0, \sigma^2)$, Lee (2007) shows that the CML estimators of β , γ , δ and σ are consistent and asymptotically efficient under regularity conditions and provided there is sufficient variation in group sizes.

Even if the assumed density of \mathbf{y}_r^* is misspecified, the pseudo CML estimator is consistent provided that the conditional mean of the \mathbf{y}_r^* s is correctly specified. This is the case since the normal density belongs to the linear exponential family (see Gouriéroux *et al.*, 1984). Of course, the estimator is no longer asymptotically efficient. Moreover, one has to compute the robust covariance matrix using the sandwich formula $J^{-1}IJ^{-1}$, where J is minus the expectation of the Hessian matrix and I the expectation of the outer-product-of-the-gradient matrix. A further advantage of this computation is that it allows us to see whether an apparent precision of CML estimators is driven by the normality assumption used in Lee (2007).

3.4.2. 2SLS and Generalized 2SLS Estimators

Alternatively, the structural equation (4) can be estimated by instrumental (IV) methods. To see how the methods work, define a $N \times N$ block-diagonal matrix $\mathbf{J} = \text{Diag}(\mathbf{J}_1, \dots, \mathbf{J}_R)$. Concatenating equation (5) over all groups yields

$$\mathbf{Jy} = \beta \mathbf{JGy} + \mathbf{JX}\gamma + \mathbf{JGX}\delta + \mathbf{J}\varepsilon \quad (6)$$

where \mathbf{y} (respectively \mathbf{X}) is obtained by stacking the vectors \mathbf{y}_r (respectively the matrices \mathbf{X}_r), for $r = 1, \dots, R$.

The reduced form of the model is

$$\mathbf{Jy} = (\mathbf{I} - \beta \mathbf{G})^{-1} (\mathbf{JX}\gamma + \mathbf{JGX}\delta) + (\mathbf{I} - \beta \mathbf{G})^{-1} \mathbf{J}\varepsilon \quad (7)$$

Identification can be given a natural interpretation in terms of instrumental variables. If $i \notin M_{r_i}$ and there are at least three different group sizes, $\mathbb{E}[\mathbf{JGy}|\mathbf{X}, \mathbf{G}]$ is not perfectly collinear to $(\mathbf{JX}, \mathbf{JGX})$ and the model is identified (see Bramoullé *et al.*, 2009, for more details). Moreover $\mathbf{JG}^2\mathbf{X}$ can be used as a matrix of valid instruments for \mathbf{JGy} .¹²

One advantage of an IV approach over pseudo CML is that it requires less structure. Specifically, we do not assume that the specified density function of the y_r s, potentially partially misspecified, is normal. Also we do not use the structure on the error terms for identification purpose. Thus identification in this case is semi-parametric, or ‘distribution-free’. Of course, this comes at a price: the IV estimator

¹¹ Note that only $n_r - 1$ elements of ε_r^* are linearly independent.

¹² 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 pre-multiplying characteristics (in deviation) by group-dependent weights and by stacking them across groups.

is asymptotically less efficient than the pseudo CML, since the latter imposes more structure on the distribution of error terms.

In addition, we can derive a generalized IV estimator as proposed in Kelejian and Prucha (1998) and discussed in Lee (2007). Assuming homoskedasticity, it yields an asymptotically optimal (best) IV estimator and reduces to a two-step estimation method in our case. More precisely, our first step consists in estimating a 2SLS as described above, by using as instruments $\mathbf{S} = (\mathbf{JX}, \mathbf{JGX}, \mathbf{JG}^2)$. The second step consists in estimating a G2SLS estimator using as instruments $\widehat{\mathbf{Z}} = (\widehat{\mathbf{JGy}}, \mathbf{JX}, \mathbf{JGX})$, where $\widehat{\mathbf{JGy}}$ is computed from the reduced form (7) pre-multiplied by \mathbf{G} and using the first-step estimates.

4. DATA

We gathered for this analysis original data from the Québec Government MERS. These administrative data provide detailed information on individual scores on standardized tests taken in June 2005 on four subjects (Mathematics, Science, 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 consists 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 has 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.¹³

There are many advantages to the use of our data. First, all fourth- and fifth-grade students must take tests on these four subjects 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 Québec. 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.

Given the lack of information on the structure of relevant social interactions, we assume that the peer 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. 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 Mathematics 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. Finally, we focus on students who take the Mathematics test in Secondary 5 (Math 514). This completes their mathematical training for secondary school. Note that the MERS administers a unique test to all secondary school students in French, History and Science. In contrast, it administers different tests in Mathematics, depending on academic options

¹³ There are more individual test scores than students as some students take tests in more than one subject area.

chosen early on by the students. We report here results for students following the regular mathematical training (Math 514). We focus on this test in our analysis.

We provide descriptive statistics in Table I. 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 Mathematics. In samples for which the regular track for the test is Secondary 5 (respectively Secondary 4), the average age of students is close to 16 (respectively 15). Most students taking French and Mathematics (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 (*Foreign variable*).¹⁴ Between 30 % and 34 % of students come from a relatively high socioeconomic background and between 40 % and 42 % from a medium one. We use an index of socioeconomic 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 socioeconomic status corresponds to the three lowest deciles of the index (high socioeconomic status to the three highest deciles).

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 Mathematics and 55,057 for History. We also observe the number of students who completed coursework but postpone test-taking to August 2005. There are 118 students postponing French, 186 postponing History, 195 postponing Science and 160 postponing Mathematics. We observe between 314 and 382 peer groups depending on the subject matter considered. The average group size is between 50 (Mathematics) and 146 (Science). The ratio between the number of groups and the average group size varies between 2.36 (French) and 7.23 (Mathematics). 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 Mathematics) 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. CML and Pseudo CML Estimates

Table II reports the results of maximum likelihood estimation with unrobust (CML) and robust (pseudo CML) standard errors. 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 . Using unrobust standard errors (in brackets), the endogenous effect is significantly different from zero and positive for Mathematics ($\hat{\beta} = 0.82$) and History ($\hat{\beta} = 0.65$). It is not significant for French ($\hat{\beta} = 0.33$) or for Science ($\hat{\beta} = -0.23$). Based on robust standard errors, it is no longer significant for History (p -value = 10.82%) but still significant for Mathematics. One thus concludes that, regarding this peer effect, inference appears to be driven by normality for one subject (History). In general, standard errors are larger using pseudo CML than CML, but their differences are not so important.

Two reasons may explain why the endogenous peer effects in Mathematics is significant in our sample. First, the standard error of the estimates is smaller in Mathematics than in other subjects. This is consistent with the fact that the average group size relative to the number of groups is

¹⁴ The language of instruction is French in most schools, and English otherwise.

Table I. Descriptive statistics

Course	Variable	Mean	SD
French (Sec. 5)	Score	72.647	14.086
	Age	16.142	0.488
	Socioecon. 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		41,778
	Number of groups		314
	Size of true groups	133.4	115.7
	Size of observed groups	133.1	115.4
Science (Sec. 4)	Score	74.689	17.671
	Age	15.255	0.610
	Socioecon. 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		54,981
	Number of groups		378
	Size of true groups	146.0	134.2
	Size of observed groups	145.5	133.7
Math ^a (Sec. 5)	Score	62.088	15.83
	Age	16.272	0.574
	Socioecon. 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		15,771
	Number of groups		361
	Size of true groups	50.7	49.9
	Size of observed groups	49.9	49.7
History (Sec. 4)	Score	70.156	17.280
	Age	15.230	0.580
	Socioecon. 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		55,057
	Number of groups		382
	Size of true groups	144.6	134.8
	Size of observed groups	144.1	134.5

^aMath^a refers to Math 514 (Secondary 5 regular course).

close to three times smaller in Mathematics than in other subjects. Second, our endogenous effect estimate is much larger in Mathematics (0.82). How does this result compare with other studies? Sacerdote (2011) has recently provided a survey of studies of endogenous peer effects in test scores for primary and secondary schools based on linear-in-means models (see his Table 4.2). Interestingly, in most reported studies (five of six) which analyze achievement in both Mathematics and Reading, the endogenous peer effect is larger in Mathematics. In addition, this effect is often very high and exceeds the value we have estimated. Thus Hoxby (2000) reports a 1.7- to 6.8-point increase in own score in relation to a 1-point increase in mean score of peers in some specifications. Betts and Zau (2004) show a 1.9-point increase in association with a 1-point increase in mean math score of peers. On the other hand, Hanushek *et al.* (2003) obtain a Mathematics peer effect

Table II. Peer effects on student achievement: conditional maximum Likelihood and pseudo conditional maximum likelihood

	French	Science	Math	History
Endogenous effect	0.296 (0.605) [0.327]	-0.231 (0.414) [0.234]	0.827** (0.319) [0.249]	0.641 (0.399) [0.272]
<i>Contextual effects</i>				
Age	-39.435** (12.798) [10.987]	-19.493* (10.237) [8.893]	0.838 (9.874) [7.382]	-31.607** (13.655) [9.471]
Socioecon. index (high)	16.613 (15.096) [17.530]	8.941 (21.637) [22.454]	29.310* (15.422) [15.580]	-6.367 (17.505) [18.947]
Socioecon. index (medium)	-4.765 (14.907) [16.870]	22.156 (18.648) [17.783]	18.246 (13.334) [13.726]	-6.713 (19.207) [18.565]
Gender (female = 1)	-24.870 (15.927) [14.393]	14.852 (13.425) [12.178]	15.558* (9.006) [9.491]	-11.837 (12.633) [12.413]
Foreign	-26.699* (14.828) [15.861]	-8.844 (13.737) [16.953]	-2.654 (12.802) [12.143]	29.148* (15.304) [18.007]
Secondary 5	167.926** (54.842) [41.179]	-0.334 (25.048) [19.956]	-6.080 (39.168) [26.056]	24.041 (24.027) [21.166]
<i>Individual effects</i>				
Age	-7.998** (0.239) [0.162]	-8.293** (0.269) [0.151]	-4.868** (0.330) [0.271]	-7.942** (0.253) [0.151]
Socioecon. index (high)	1.423** (0.308) [0.245]	1.609** (0.297) [0.268]	2.112** (0.496) [0.500]	2.019** (0.322) [0.261]
Socioecon. index (medium)	0.670** (0.266) [0.220]	0.785** (0.260) [0.230]	1.189** (0.464) [0.435]	0.795** (0.272) [0.234]
Gender (female = 1)	3.807** (0.196) [0.162]	0.319 (0.200) [0.158]	1.018** (0.325) [0.301]	-1.641** (0.207) [0.159]
Foreign	-2.596** (0.314) [0.279]	2.095** (0.380) [0.278]	-0.081 (0.513) [0.548]	0.806** (0.384) [0.284]
Secondary 5	10.519** (1.258) [0.676]	1.653** (0.560) [0.328]	6.474** (1.096) [0.767]	3.126** (0.537) [0.399]
Log-likelihood	-162548.552	-226078.181	-62420.961	-226216.108

Notes:

CML unrobust standard errors in brackets; pseudo CML robust standard errors in parentheses.

Asterisks indicate significance level at **5%; *10%; based on robust SE.

The dependent variable is the score in June 2005 provincial secondary exams.

of 0.4.¹⁵ Thus our estimate lies on the average to high side of the range of previous estimates. Observe finally that our results in Mathematics are larger than those usually obtained in studies based on randomized experiments (e.g. Sacerdote, 2001; Zimmerman, 2003). One possible explanation is that peers used in these papers are often people from the same dorm. These individuals do not necessarily represent those who exercise significant influence on students' scholarly achievement.

The relatively large endogenous peer effect in Mathematics may reflect the fact that mathematics provides more opportunities for interactions among students. Also, probably more than in other subjects, it may also reflect general effects such as disruption. For instance, it is

¹⁵ Kang (2007, p. 475) also provides a survey of endogenous peer effects in achievement in Mathematics which is broadly consistent with results reported in Sacerdote (2011).

likely that success in Mathematics requires much concentration in class from the average student. Now suppose that there is a student (with a low grade in Mathematics) in class who is characterized by his propensity to disrupt learning by bad behavior or asking poor questions. His behavior may have large negative effects on his peers' scholarly achievement (see, for example, Lazear, 2001) and thus generates strong endogenous peer effects.

Regarding individual characteristics, most of them 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. For instance, results from the 2000 Program for International Student Assessment (PISA) show that Québec 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 Québec Government, 2001). Similarly, in our analysis, the difference in performance is quantitatively large in French but much smaller in the other disciplines. The performance of foreign students is, not surprisingly, significantly lower than for non-foreign students on the French test, but higher for Science and History and not significantly different for Mathematics. 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.

As far as contextual variables are concerned, a few of them have a significant impact on student performance. Average age of other students has a negative and significant effect on all test scores except Mathematics, where it is positive but not significant. These results also conform our expectations. When the number of repeaters rises (as reflected by an increase in mean age of our peers at a given grade level), this will tend to reduce own test score. The proportion of other students enrolled in Secondary 5 has a large positive and significant effect on own score in French. Peers' socioeconomic background has little effect on own schooling performance. The proportion of female students among peers has a positive and significant effect in Mathematics. 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 in the characteristics of *every* other student in the group.¹⁶

5.2. Reflection Problem

One 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 II but excluding contextual effects that are not individually significant in the pseudo CML specification (i.e. for which the null that $\delta = 0$ is not rejected); see Table 1 of the supplementary data Appendix, available online as supporting information. Using likelihood ratio tests, we reject the null that these δ s are jointly equal to zero for French but not for the other subjects. This suggests that the exclusion restrictions may be valid for these latter samples. Therefore, the pseudo CML estimators provided in Table 1 of the supplementary Appendix should be consistent and asymptotically more efficient than those provided in Table II of the main text for the Science, Mathematics and History tests. Results, however, appear to be robust to these new specifications. Observe finally 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 the different types of peer effects at work.

¹⁶ We have also estimated a second-order pseudo CML in which restrictions are directly incorporated into the variance term and estimated. Results are quite similar to those presented in Table II.

5.3. 2SLS and G2SLS Estimates

Table III provides the 2SLS 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 and pseudo CML estimates of Table II, none of the endogenous effects is 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 III also reports two standard test results giving information on instrumental variables properties. We first look at Sargan tests on the validity of instruments and the over-identification restrictions of the model. We do not reject the null for Science, Mathematics and History, but we reject it for French. While this may indicate a problem of model specification in this last case, one must be cautious in interpreting the test given the likely low convergence of peer effects IV estimates. We then compute Stock and Yogo test statistics on weak identification. 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. Finally, Hausman tests have been performed to test the equality of pseudo CML and G2SLS estimators. Under the null, both of these estimators are consistent, but pseudo CML estimators are asymptotically more efficient; under the alternative, G2SLS estimators are consistent whereas pseudo CML estimators are not. For each subject, we could not reject the null. This suggests the absence of specification errors in the model. Finally G2SLS estimates are provided in Table 2 of the supplementary data Appendix. Results are qualitatively similar to those of Table III.

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 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 (see Table I), and we suspect that this may help identification. To study these issues, we realize two simulation exercises. First, we vary group sizes in a systematic manner and study how this affects the bias and precision of estimators. To focus on the approach which provides the most reasonable findings in our empirical analysis, we report results on the model using CML.¹⁷ We look at uniform distributions, vary the size of the distribution's support 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 dispersion in group sizes on the strength of identification. In particular, conditional maximum likelihood performs well on fully calibrated simulations. This suggests that the bias due to small sample issues is likely low in the results presented in Table II.

For each simulation exercise, we keep the number of observations fixed around 42,000 and run 1000 replications. We first consider average sizes of 10, 20, 40, 80 and 120. We pick group sizes from the following intervals with decreasing length:

¹⁷ In an earlier version of the paper, we also provided results for IV estimates. Basically, the results are qualitatively the same for IV as those for CML but, as expected, the magnitude of the bias and the loss in precision are always larger for IV than for CML.

Table III. Peer effects on student achievement: 2SLS estimation with group fixed effect

	French	Science	Math	History
Endogenous effect	1.378 (1.468)	-0.509 (0.764)	-0.037 (0.477)	0.787 (0.980)
<i>Individual effects</i>				
Age	-7.690** (0.197)	-7.962** (0.167)	-4.606** (0.228)	-7.609** (0.163)
Socioecon. index (high)	1.373** (0.242)	1.754** (0.250)	1.836** (0.423)	2.041** (0.248)
Socioecon. index (medium)	0.661** (0.221)	0.826** (0.219)	1.069** (0.365)	0.803** (0.221)
Gender (female = 1)	3.871** (0.164)	0.333** (0.159)	0.965** (0.265)	-1.553** (0.157)
Foreign	-2.514** (0.282)	2.128** (0.270)	-0.005 (0.496)	0.716** (0.276)
Secondary 5	9.516** (0.781)	1.415** (0.327)	6.674** (0.741)	2.910** (0.390)
<i>Contextual effects</i>				
Age	4.205 (4.845)	13.496** (3.050)	6.713** (1.712)	8.552** (4.036)
Socioecon. index (high)	7.364 (17.305)	30.997* (16.678)	15.962** (7.641)	-6.246 (15.620)
Socioecon. index (medium)	-7.103 (16.813)	26.344* (13.908)	13.501* (7.555)	-8.047 (14.598)
Gender (female = 1)	-21.310* (12.261)	15.637 (12.202)	13.237** (5.808)	0.567 (11.708)
Foreign	-15.732 (12.571)	-2.232 (11.449)	-0.065 (7.189)	19.385 (12.903)
Secondary 5	40.184 (36.380)	-17.370 (14.470)	7.825 (21.360)	2.537 (23.060)
Sargan test	23.52	0.54	1.40	5.35
[<i>p</i> -value]	[0.00]	[1.00]	[0.97]	[0.50]
Stock and Yogo test	706.84	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]

Notes:

Robust standard errors in parentheses.

Asterisks indicate significance level at **5%; *10%.

The dependent variable is the score in June 2005 provincial secondary exams.

- 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:

- pick a group size from a uniform distribution for which the support is defined by the minimum and maximum value of the interval;
- truncate this value by eliminating its decimal portion;
- repeat steps 1 and 2 as long as the total number of observations is below or equal to 42,000.

To reduce computing time, we assume that 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 close to the estimated coefficients for the French test: $\beta = 0.35$, $\gamma_{\text{age}} = -8$, $\gamma_{\text{gender}} = 3.8$, $\delta_{\text{age}} = -40$, $\delta_{\text{gender}} = -25$.

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.

Looking at Table IV, we first compare simulation results across average group sizes and then we examine how estimators perform for a given average group size as dispersion in group size decreases. Separate horizontal panels in Table IV pertain to different values of average group size. We report the average estimated coefficient and standard error for the endogenous effect (first vertical panel), the contextual effect associated with age (second vertical panel) and the contextual effect associated with gender (third vertical panel). We find that even for the largest average group size (i.e. 120), CML may perform well in terms of bias and precision (first line in the last horizontal panel of Table IV). The biases of CML get larger, in general, as average group size increases. The CML estimate of the endogenous effect attains a plateau at the value 1. This is consistent with the fact that the CML estimator tends towards the naive OLS estimator as group sizes become larger. In general, peer effects are also less precisely estimated in large groups than in small groups.

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 CML typically increases, while the precision typically decreases. In Table IV, this amounts to looking at each horizontal panel separately. Observe, however, that since we roughly pick group sizes from a uniform distribution holding average group size fixed, reducing the interval's length affects the two parameters of the size distribution (i.e. the minimum and maximum value of its support) and a number of its moments. In particular, this leads to a reduction in variance and to an increase in the size of the smallest groups. In general, both the variance and the size of smallest groups may matter and the strength of identification may depend on the size distribution in complex ways. We leave a deeper investigation of this issue to future research.

Table IV. Group size variation: simulations using CML

Avg. group size	Group sizes Range	Endogenous effect		Contextual effects: age		Contextual effects: gender	
		CML		CML		CML	
		Avg. coeff.	SE	Avg. coeff.	SE	Avg. coeff.	SE
10	[3; 17]	0.35	0.00	-40.01	0.25	-25.01	0.33
10	[5; 15]	0.35	0.00	-40.00	0.35	-24.99	0.74
10	[7; 13]	0.35	0.02	-40.00	0.53	-25.01	1.50
10	[9; 11]	0.57	0.38	-40.27	1.79	-26.97	5.78
20	[3; 37]	0.35	0.00	-40.01	0.27	-25.03	0.44
20	[8; 32]	0.35	0.02	-40.00	0.50	-25.02	1.10
20	[13; 27]	0.35	0.09	-39.95	1.23	-25.04	2.11
20	[18; 22]	0.94	1.56	-37.98	5.37	-28.55	8.47
40	[3; 77]	0.35	0.00	-39.98	0.41	-25.03	0.65
40	[12; 68]	0.36	0.07	-39.97	1.42	-25.05	1.66
40	[21; 59]	0.39	0.20	-39.85	2.76	-25.14	2.67
40	[30; 50]	0.72	0.85	-37.92	5.82	-26.93	5.30
40	[39; 41]	1.00	155.98	-36.25	78.67	-26.28	69.22
80	[3-157]	0.35	0.01	-39.99	0.68	-25.05	0.98
80	[18-142]	0.43	0.19	-39.55	2.93	-25.40	2.49
80	[33-127]	0.57	0.46	-38.54	4.87	-25.96	3.66
80	[48-112]	0.87	1.20	-36.47	8.02	-27.10	5.75
80	[63-97]	1.00	5.27	-35.75	17.05	-27.74	11.97
120	[3-237]	0.36	0.01	-39.99	0.99	-25.10	1.50
120	[28-212]	0.64	0.51	-38.14	5.22	-26.34	3.79
120	[53-187]	0.89	1.30	-35.99	8.69	-27.25	5.85
120	[78-162]	1.00	3.85	-35.34	15.16	-27.65	9.94
120	[103-137]	1.00	25.15	-35.32	39.00	-28.20	25.29

Note: True value of parameters: endogenous effect, 0.35; contextual effects: age, -40; contextual effects: gender, -25.

We next fully calibrate the simulations' parameters on the data. We use observed group sizes in the French sample, calibrate the model parameters $\{\beta, \gamma_{\text{age}}, \gamma_{\text{gender}}, \delta_{\text{age}}, \delta_{\text{gender}}\}$ and moments of the explanatory variables as previously, and set the variance of the error term in the structural equation equal to the estimated variance in the French sample ($\hat{\sigma}^2 = 154.7$). Simulation results which now report both CML and IV estimates are reported in Table V. 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 for CML what we obtained from picking group sizes at random; they show that dispersion in group sizes helps identification. Besides, this suggests that small sample bias may be relatively high in the IV estimates of Tables III, and of Table II of the supplementary Appendix, but relatively low for the CML estimates of Table II.

7. CONCLUSION

This paper provides an analysis of social interactions in scholar achievement when students interact through groups. Based on a linear-in-means approach with group fixed effects (Lee, 2007), we make two main contributions regarding the identification and estimation of peer effects. First, we provide a new intuition for identification. We show that full identification of the model relies on three key properties. (i) Since the individual is excluded from his peer group, above-average students have below-average peers (with respect to any attribute). Therefore, when individual and peer effects are positive, peer effects then tend to reduce the dispersion in outcomes. (ii) This reduction is stronger in smaller groups, reflecting the larger effect of excluding one individual from the mean. (iii) Contextual and endogenous peer effects generate reductions of different shapes, which allow us to identify both of them.

Second, as regards the estimation of peer effects, 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 Québec (Canada). Based on a pseudo conditional maximum likelihood approach, our results indicate that students significantly benefit from their peers' higher test scores in Mathematics but not in other subjects such as Science, History and French. Two reasons may explain these results. First, this is likely to reflect the fact that Mathematics provides more opportunities for interactions among students. Second, in our sample, the average group size (relative to the number of groups) is close to three times smaller in Mathematics than in other subjects.

Table V. Simulations calibrated on French sample (1000 replications)

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

Notes:

Average standard errors are in parentheses. Group sizes are calibrated on our French sample. $\sigma^2 = \hat{\sigma}^2$ (calibrated) = 154.704. True value of parameters: endogenous effect, 0.35; individual effects: age: -8; individual effects: gender, 3.8; contextual effects: age: -40; contextual effects: gender: -25.

As suggested by Lee (2007), accurate estimation of peer effects requires relatively small groups. This is also confirmed by our Monte Carlo simulations. These results should be warning applied researchers in the future against using data in which the size of groups is too large. Besides, our simulations indicate that, for a given average group size, increasing group size dispersion improves the precision of peer effects estimates. In fact, our results suggest that, conditional on estimating on the whole sample, even data on larger groups may provide useful information for estimation purposes. The basic intuition is that data on very large groups can be used to provide more precise individual effects estimators. In turn, this indirectly provides more efficient estimates of the peer effects from data on smaller groups. Thus future estimations of Lee's model may benefit from data with relatively small average group size but relatively large group size dispersion, including both small and large groups.

In terms of public policy, the fact that the endogenous peer effects appear to be very large in Mathematics suggests that a reform that improves the amount and quality of Mathematics learning is likely to yield very high returns in terms of scholar achievement. This is so since such a reform will not only have direct effects on student performance in Mathematics but also strong indirect effects through the additional external benefits generated by the social multiplier. Remarkably, our analysis also shows that the indirect peer effects of the reform will reduce performance inequalities in Mathematics across students. This is the case because low-ability students have better peers (since their peers exclude them) and high-ability students have worse peers (for the same reason). Moreover, the strong negative effects of the average age of peers on scholar achievement (except in Mathematics) suggest that resources invested by the government to reduce the number of repeaters may have an important indirect positive impact on student performance. One limitation of Lee's linear-in-means approach is that it imposes that average test scores over all schools are not influenced by a reallocation of students across schools (see Sacerdote, 2011). Therefore, the model does not have much to say about issues such as optimal school composition by race or ability.

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 heterogeneous (as in Sacerdote, 2001). One could also 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 (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 a nonlinear version of Lee's approach. Thus 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 for public policy.

ACKNOWLEDGEMENTS

We thank Lung-fei Lee, seminar participants at the University of Toronto, Northwestern University, Université de Paris 1, and the CIRANO-CIREQ Conference on the Econometrics of interactions for insightful discussions, and three anonymous referees and the co-editor Thierry Magnac for very helpful comments. We are also grateful to the Québec Ministry of Education, Recreation and Sports (MERS) for providing the data, in particular Raymond Ouellette and Jeannette Ratté for their assistance in obtaining and interpreting the data used in this study. The views expressed in this paper are solely our own and do not necessarily reflect the opinions of the MERS. We received excellent research assistance from Steeve Marchand. Support for this work has been provided by the Canada Chair of Research in Economics of Social Policies and Human Resources, and le Fonds Québécois de Recherche sur la Société et la Culture and le Centre Interuniversitaire sur le Risque, les Politiques Économiques et l'Emploi.

REFERENCES

- 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.
- Angrist JD, Lang K. 2004. Does school integration generate peer effects? Evidence from Boston's Metco Program. *American Economic Review* **94**(5): 1613–1634.
- Betts JR, Zau A. 2004. Peer groups and academic achievement: panel evidence from administrative data. Working paper, University of California, San Diego, and Public Policy Institute of California.
- Bifulco R, Fletcher JM, Ross SL. 2011. The effect of classmate characteristics on post-secondary outcomes: evidence from the Add Health. *American Economic Journal: Economic Policy* **3**(1): 25–53.
- Bramoullé Y, Djebbari H, Fortin B. 2009. Identification of peer effects through social networks. *Journal of Econometrics* **150**: 41–55.
- Brock W, Durlauf S. 2007. Identification of binary choice models with social interactions. *Journal of Econometrics* **140**: 57–75.
- Calvó-Armengol A, Patacchini E, Zenou Y. 2009. Peer effects and social networks in education. *Review of Economic Studies* **76**(4): 1239–1267.
- Davezie L, d'Haultfoeuille X, Fougère D. 2009. Identification of peer effects using group size variation. *The Econometrics Journal* **12**: 397–413.
- Epple D, Romano RE. 2011. Peer effects in education: a survey of the theory and evidence. In *Handbook of Social Economics*, Benhabib J, Bisin A, Jackson MO (eds). North-Holland: Amsterdam; 1053–1163.
- Evans W, Oates W, Schwab R. 1992. Measuring peer group effects: a study of teenage behavior. *Journal of Political Economy* **100**(5): 966–991.
- Gaviria A, Raphael S. 2001. School based peer effects and juvenile behavior. *The Review of Economics and Statistics* **83**(2): 257–268.
- Gouriéroux C, Montfort A, Trognon A. 1984. Pseudo maximum likelihood methods: theory. *Econometrica* **52**: 681–700.
- Graham BS. 2008. Identifying social interactions through conditional variance restrictions. *Econometrica* **76**(3): 643–660.
- Hanushek E, Kain J, Markman J, Rivkin S. 2003. Does peer ability affect student achievement?. *Journal of Applied Econometrics* **18**(5): 527–544.
- Hoxby C. 2000. Peer effects in the classroom: learning from gender and race variation. Working paper no. 7867, National Bureau of Economic Research, Cambridge, MA.
- Kang C. 2007. Classroom peer effects and academic achievement: quasi-randomization evidence from South Korea. *Journal of Urban Economics* **61**: 458–495.
- Kelejian HH, Prucha IR. 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* **17**: 99–121.
- Krueger AB. 2003. Economic considerations and class size. *The Economic Journal* **113**(485): F34–F63.
- Lazear EP. 2001. Educational production. *Quarterly Journal of Economics* **116**: 777–803.
- Lee LF. 2007. Identification and estimation of econometric models with group interactions, contextual factors and fixed effects. *Journal of Econometrics* **140**(2): 333–374.
- Lin X. 2010. Identifying peer effects in student academic achievement by spatial autoregressive models with group unobservables. *Journal of Labor Economics* **28**(4): 825–860.
- Manski C. 1993. Identification of endogenous social effects: the reflection problem. *Review of Economic Studies* **60**(3): 531–542.
- Moffitt R. 2001. Policy interventions, low-level equilibria, and social interactions. In *Social Dynamics*, Durlauf S, Young P. MIT Press: Cambridge, MA; 45–92.
- Québec Government. 2001. PISA 2000: The Performance of Canadian Youth in Reading, Mathematics and Science. Results for Québec Students Aged 15. Ministry of Education, Recreation and Sports.
- Rivkin SG. 2001. Tiebout sorting, aggregation and the estimation of peer group effects. *Economics of Education Review* **20**: 201–209.
- Sacerdote B. 2001. Peer effects with random assignment: results for Dartmouth roommates. *Quarterly Journal of Economics* **116**(2): 681–704.
- Sacerdote B. 2011. Peer effects in education: how might they work, how big are they and how much do we know thus far? In *Handbook of the Economics of Education*, Vol. **3**, Hanushek EA, Machin S, Woessmann L (eds). North-Holland: Amsterdam; 250–277.
- Stinebrickner R, Stinebrickner TR. 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* **90**: 1435–1454.
- Zimmerman D. 2003. Peer effects in academic outcomes: evidence from a natural experiment. *The Review of Economics and Statistics* **85**(1): 9–23.