DISCUSSION PAPER SERIES

No. 5682

CLOSE NEIGHBOURS MATTER: NEIGHBOURHOOD EFFECTS ON EARLY PERFORMANCE AT SCHOOL

Dominique Goux and Eric Maurin

LABOUR ECONOMICS

Centre for Economic Policy Research

www.cepr.org

Available online at: www.cepr.org/pubs/dps/DP5682.asp

ISSN 0265-8003

CLOSE NEIGHBOURS MATTER: NEIGHBOURHOOD EFFECTS ON EARLY PERFORMANCE AT SCHOOL

Dominique Goux, INSEE, Paris
Eric Maurin, Paris Jourdan Sciences Economiques and CEPR

Discussion Paper No. 5682
May 2006

Centre for Economic Policy Research
90–98 Goswell Rd, London EC1V 7RR, UK
Tel: (44 20) 7878 2900, Fax: (44 20) 7878 2999
Email: cepr@cepr.org, Website: www.cepr.org

CEPR Discussion Paper No. 5682

May 2006

ABSTRACT

Close Neighbours Matter: Neighbourhood Effects on Early Performance at School*

Children's outcomes are strongly correlated with those of their neighbours. The extent to which this is causal is the subject of an extensive literature. An identification problem exists because people with similar characteristics are observed to live in close proximity. Another major difficulty is that neighbourhoods measured in available data are often considerably larger than those which matter for outcomes (i.e. close neighbours). Several institutional features of France enable us to address these problems. We find that an adolescent's performance at the end of junior high-school are strongly influenced by the performance of other adolescents in the neighbourhood.

JEL Classification: I21 and J24
Keywords: neighbourhood effects on education

Dominique Goux
INSEE
15, Bd Gabriel Péri
BP 100, Timbre G231
92244 Malakoff Cédex
FRANCE
Tel: (33 1) 41 17 5442
Fax: (33 1) 41 17 6163
Email: dominique.goux@insee.fr

For further Discussion Papers by this author see:
www.cepr.org/pubs/new-dps/dplist.asp?authorid=131059

Eric Maurin
Ecole Normale Supérieure
Bureau 105, Bat B
48 Bd Jourdan
75014 Paris
FRANCE
Tel: (33 1) 43 13 6263
Email: Eric.Maurin@ens.fr

For further Discussion Papers by this author see:
www.cepr.org/pubs/new-dps/dplist.asp?authorid=131502

* Previous versions of this paper were presented at the Conference in memory of Zvi Griliches (Paris) and at the CEPR Conference on Changing Condition in Education (Uppsala), at the university of Pompeu Fabra (Barcelone), at the university of Milan, at the CREST (Paris) and at the CEPR symposium on Public Policy (Paris). This research has benefited from a grant of the French Commissariat Général au Plan (n_ 7-2002). We thank Esther Duflo, Steve Gibbons, Per Johansson, Thomas Piketty, Gerard Van den Berg, and Louis-André Vallet for their comments on earlier versions of this paper. Special thanks are due to Sandra McNally for her insightful comments and very helpful suggestions for improving the paper.

Submitted 18 April 2006

1 Introduction

The assumption that children's outcomes are influenced by the characteristics and outcomes of their neighbours forms the basis of a large and growing literature in the social sciences. Providing a convincing evaluation of neighbourhood effects has proven very difficult, however. The main difficulty is in isolating variation in neighbourhood attributes which are exogenous to children's outcomes. Children and families living in the same neighbourhood tend to have similar outcomes. It is unclear, however, whether this is because they influence each other or because they share the same unobserved characteristics (see Manski, 1993, Ginther et al., 2000, Moffitt, 2001, Brock and Durlauf, 2001). Another issue is that neighbourhoods measured in available data sets are often considerably larger than those which matter for outcomes (i.e. close neighbours).

This paper addresses these issues and identifies the causal impact of close neighbours' characteristics on children's outcomes. The French Labour Force Survey enables us to consider the effect of close neighbours because of the nature of data collection: the basic sampling unit consists of groups of 20 to 30 adjacent households. It provides us with a large sample of 15 year-old adolescents and includes detailed information on the situation of all the other adolescents and adults living in the close neighbourhood, defined as the 20 to 30 adjacent houses. Existing studies typically proxy neighbourhoods with census tracts - relatively large groups of people (several thousand). Assuming that distant neighbours have less influence than close ones, this plausibly leads to an underestimate of the influence of close neighbours. Our data provide an interesting opportunity to overcome this difficulty and to analyse how persons living in adjacent houses actually influence each other[1].

Our first identification strategy relies on variation across neighbourhoods in the proportion of adolescents born at the beginning (or at the end) of the year. As discussed below, the date of birth within the year is an important determinant of French children's early performance at school and is plausibly exogenous to the quality of the neighbourhood in which they live. In such a context, one simple way to identify contextual effects is to test whether children's performance at school are affected by the distribution of dates of birth within the year of the *other* children living in the same neighbourhood. As shown below, the answer is positive. Regardless of their own date of birth, children living in a neighbourhood with a relatively high proportion of children born at the beginning of the year perform significantly better than children living in a neighbourhood with a relatively high proportion of children born at the end of the year. This result provides interesting evidence of contextual effects on children's outcomes. It is possible to take some steps further by focusing on children's *late* outcomes - at the end of junior high-school - and by assuming that their neighbours' dates of birth, as such, have no effect on these outcomes, i.e. neighbours' date of birth, as such, has no influence on outcomes, except maybe

[1] There is a related literature which studies interactions among close neighbors, although the focus is not on educational attainment (see, for example, Loannides, 2002, Loannides, 2003, Lonnadies and Zabel, 2003, Case and Katz, 1991)

on early outcomes in primary school or at the beginning of junior high-school. In such a case, the distribution of neighbours' dates of birth can be used as an instrumental variable to identify the effect of neighbours' early educational advancement on an adolescent's performance at school. Our IV estimates suggest that a one SD increase in the proportion of neighbours who have already been held back a grade at age 15 increases an adolescent's probability of grade repetition between the age of 15 and 16 by about 10-15 percentage points (i.e. about 20% of a SD). It is shown that these estimates do not depend on the specific characterisation of neighbours' date of birth which is used for idenfication. Overidentification tests do not reject the identifying assumption nor the linear specification of the endogenous effect which is used in this paper.

Following the terminology introduced by Manski (1993), our first strategy identifies the endogenous effect, i.e. the effect of neighbours' outcomes on own outcomes. We have developed a second strategy which provides an evaluation of the reduced-form effect of neighbours' family background in relatively poor neighbourhoods. This approach relies on available information on families living in public housing (*Habitation à Loyer Modéré*, hereafter HLM, about 20% of the population). In France, any family is eligible for a HLM provided that the income per unit of consumption is sufficiently low. The problem is that the number of eligible families is about three times as large as the available space in HLM. Rents are also considerably lower in HLM than in non-HLM. Given these facts, the turnover is very low and the waiting lists are very long. HLM managers have a very limited set of housing to offer each year to eligible families and very little control over the specific neighbourhoods to which families can be assigned. We provide various specification checks supporting the assumption that HLM assignment is quasi-random. Under this assumption, HLM neighbourhood membership can be considered as exogenous and neighbourhood effects can be identified through standard regressions. Interestingly, they confirm the existence of strong contextual effects in HLM neighbourhoods.

The French Labor Force Survey provides information on adolescents' outcomes only. To further explore the influence of social context on French children, we have used a longitudinal survey recently conducted by the French Ministry of Education. It provides detailed information on the early school career of a large representative sample of pupils. This dataset makes it possible to analyze the relationship between the scores in national tests at entry into 3rd grade and the characteristics of peers at entry into 1^{st} grade, using exactly the same reduced-form and IV specifications as with the Labor Force Survey. Most interestingly, the reduced-form analysis confirms that individual scores obtained at entry into third grade decrease significantly with the proportion of first-grade peers who were born at the end of the year. Also, IV estimates suggest that a one SD increase in the average score of early peers leads to an increase of about 30% of a SD of a child's score at entry into the third grade. As it turns out, the influence of early peers does not seem less strong than that of neighbours with whom we interact later in life.

Generally speaking, this paper contributes to the literature on the influence of peers on own educational achievement, where peers are defined as children of a similar age and likely to interact. There is no consensus on the importance of peer effects on own achievement in this literature. Some papers report significant effects (e.g. Ding and Lehrer, 2004, Hoxby, 2001, or McEwan, 2003) whereas others find no impact at all (e.g. Angrist and Lang, 2004). One explanation for the lack of consensus is variation in how 'neighborhood' is defined, as well as the variety of approaches used to identify peer effects. A branch of this literature identifies peers' influence through the analysis of housing mobility programs where some low income inner-city families are given assistance in moving to less segregated, randomly selected locations (e.g. Jacob, 2004, Leventhal and Brooks-Gunn, 2004, Liebman et al. 2004, Ludwig et al., 2001, Ludwig et al. 2000, Rosembaum, 1995, Sanbonmatsu et al. 2004). Another branch of the literature identifies contextual effects in the classroom or at university through exploiting either random variation in the peer group composition over time or random assignment of peers to individual students (e.g. Gibbons and Telhaj, 2005, Gould et al., 2004, Hanushek et al., 2003, Kremer and Levy, 2003, Pischke and Ammermueller, 2006, Sacerdote, 2001, Zimmerman, 2003). Using sibling data, Aaronson (1998) finds a significant correlation between variation in educational outcomes across siblings and variation in neighbourhood quality arising from family change of residence. Other researchers have tried to address endogeneity issues by developing instrumental variables strategies (e.g. Cutler and Glaeser, 1997; Evans et al., 1992). For example, Cutler and Glaeser (1997) use the topographical features of cities to identify the effect of spatial segregation on blacks' outcomes.

This paper is organized as follows. The next section provides a description of our data. Section 3 shows the results of the strategy using neighbours' dates of birth within the year as an instrumental variable. Section 4 shows results of the strategies which build on the available information on public sector housing. Section 5 shows the findings with the panel of pupils. Section 6 concludes.

2 Data and Variables

The datasets used in this paper come from 12 waves of the French Labor Force Survey (LFS), conducted each year between 1991 and 2002. One interesting feature of the French LFS is that the basic sampling unit consists of groups of about 20 adjacent households (*aires*). More specifically, a typical LFS consists of a representative sample of about 3,500 *aires*. We take "neighbourhood" as equivalent to the LFS *aire*. Each year, within each *aire*, all the households are surveyed and, within each household, all persons aged 15 or more are interviewed. The French statistical office (INSEE) has chosen this sampling strategy so as to reduce the travelling expenses of those who administer the survey.

For each respondent, we have standard information on his date of birth, sex, nationality, family situation, place of birth, education, labor market situation (unemployed, out of the labor force, employed). Also we know whether the respondent has been living in his current residence for one year or whether he has just moved into the neighbourhood. For respondents who are still in the education system, we know their current grade. By comparing their age and grade, we know whether they have been held back a grade in primary or junior high-school. For example, respondents of year t born in $t-15$ are in the ninth grade (at least) if they have not been held back a grade[2]. In the French context, to repeat a grade in elementary school or junior-high school is a very direct indicator of early performance at school. The recent Program for International Student Assessment (PISA) conducted by the OECD shows that 15 year-old French adolescents who have repeated a grade obtain much lower scores in mathematics, reading or science than normal-age adolescents. The difference is about 1.14 standard deviations of the score in mathematics, 1.26 SD in reading, 1.17 SD in science (see Murat and Rocher, 2003). By the end of junior high-school, about 42% of adolescents have been held back a grade.

Another interesting feature of the French Labor Force Survey is that only one-third of the sample is renewed each year. For each t, we can construct a large representative sample of 15 year-old adolescents with information on their situation at t and at $t+1$ ($N = 13,100$). This paper will focus on the sample of 15 year-old respondents who were already living in their house one year before, who are still observed in the LFS at $t+1$ and such that we observe at least one other 15 year-old adolescent in their *aire*. Tables A1 and A2 in Appendix A provide the distribution of the adolescents in our sample according to the number of other 15 year-old adolescents observed in their *aires* and provide the basic descriptive statistics for these adolescents.

For each adolescent, it is possible to calculate the proportion of other adolescents in the *aire* who have been held back a grade in primary or junior high-school and who are not 'normal-age'. The basic research question is whether an adolescent's probability of repeating a grade between the age of 15 and 16 is affected by the proportion of 15 year-old neighbours who are not normal-age. Does the variation in an adolescent's educational advancement between the age of 15 and 16 depend on the educational advancement of his/her close neighbours of the same age?

For each adolescent, we have also constructed several explanatory variables describing the average characteristics of other families living in the *aire*, namely the proportion of single-parent families, the proportion of families with 3 or more children, the proportion of non-French or unemployed workers among the adults in these families, the proportion of high-school dropouts, the proportion of college graduates. Using the terminology of Manski (1993), the impact of these variables on an adolescent's educational advancement corresponds to exogenous effects. Let us emphasize that for each respondent the different *aire*-level indicators are constructed using only information on individuals who do not belong to the family of the respondent[3].

[2] Put differently, 15-years-old respondents who are in the ninth grade are "normal-age", i.e., the expected age for entry to a given grade without repeating or skipping any grades up to that time.

[3] For example, the proportion of adolescents in the neighborhood born at the beginning (or at the end) of the year is constructed without using the date-of-birth of the respondent. This only uses the date of birth of other adolescents living in the aire.

3 Identification using information on neighbours' months of birth

The question is whether an adolescent's educational advancement at the end of junior high-school is affected by the characteristics of other adolescents in the neighbourhood. The first identification strategy builds on the use of a variable which determines children's performance at school, but which is nonetheless exogenous to the quality of the neighbourhood in which they live. Specifically, our first approach relies on the fact that date of birth within the year is an important determinant of French children's early performance at school and that this is plausibly exogenous to the quality of their neighbourhood.

There is ample evidence showing that French children's date of birth within the year is an important determinant of their early school outcomes[4]. The French school system is characterized by the full day character of both pre-primary and primary school, the heavy teaching load, and the very high proportion of pupils who have to repeat one or two grades before the end of compulsory schooling. In such a context, the date of birth within the year is a very important determinant of early school performance - plausibly more important than in most other Western countries. The national evaluations conducted each year at entry into third grade show an average difference of about 1/2 of a standard deviation between the scores of children born in January and those of children born in December. The proportion of 15-years-old children held back a grade is about 15 percentage points higher for children born at the end of the year - the least mature of their class - than for children born at the beginning.

In contrast, there is no strong reason for children's date of birth within the year to be correlated with the quality of the neighbourhood in which they live. As discussed below, there is no specific residential concentration of children born at the beginning (or the end) of the year. In such a context, it is possible to develop a very simple test for the existence of contextual effects.

To better understand why this is the case, let us denote y_k children's educational advancement at age k ($k = 7, ..., 16$) and let us assume that y_k is defined recursively by,

$$y_k = \alpha_{k-1} E(y_{k-1} \mid n) + \beta_{k-1} E(v \mid n) + \gamma_{k-1} v + \theta_{k-1} y_{k-1} + \epsilon_n + u, \qquad (1)$$

where v represents date of birth within the year, n the neighbourhood, ϵ_n a neighbourhood fixed effect (the quality of schools) and u the omitted individual characteristics, i.e. the resources that affect schooling and that an individual can bring from one neighbourhood to another. The γ_k parameters represent the effect of own date of birth within the year on own performance at school whereas

[4] There is a literature on the influence of date of birth within the year (called the relative age effect) going back a few decades (see, for example, Barnsley et al. 1985, Allen and Barnsley, 1993). It is shown that relative age has an impact on achievement in competitive activities (Hockey, Soccer), on achievement at school, on emotional development and even on the probability of committing suicide.

the θ_k parameters capture the persistance of educational outcomes over time[5]. Using Manski's terminology, the α_k parameters represent endogenous effects, whereas the β_k parameter captures an exogenous contextual effect.

The omitted resources u are likely to be correlated with ϵ_n, but the date of birth of the respondents (v) and the distribution of the date of birth of their neighbours ($E(v \mid n)$) are plausibly uncorrelated with the other determinants of performance at school, as defined by ϵ_n or u. Tables A3 and A4 in the appendix show that there is no significant correlation between the basic observable determinants of an adolescent's performance at school (i.e. date of birth, gender, nationality or family background) and the proportion of neighbours born at the beginning (or at the end) of the year. In particular, there is no specific correlation between own date of birth and neighbours' dates of birth. There are good and bad neighbourhoods (i.e., high and low ϵ_n), but children and their neighbours do not seem to be sorted across good and bad neighbourhoods according to their date of birth.

After averaging Equation (1) conditional on n and solving the recursive system of equations, we obtain a first-stage equation that can be written,

$$E(y_k \mid n) = \beta_{1,k-1} E(v \mid n) + \omega_n, \tag{2}$$

where the new fixed effect ω_n is a linear combination of ϵ_n and $u_n = E(u \mid n)$ whereas the new parameter $\beta_{1,k-1}$ is a linear combination of the $\beta_{k-t} + \theta_{k-t}$ parameters, $t = 1, ..., k-6$ (see Technical Appendix). The $\beta_{1,k-1}$ parameter captures the cumulative impact of the average maturity of children living in a neighbourhood on their average outcome at age k. This equation makes it clear why the different $E(y_k \mid n)$ are likely to be correlated with ϵ_n and, consequently, why α_{k-1} cannot be estimated in Equation 1 through a standard linear regression of y_k on $E(y_{k-1} \mid n)$.

Replacing $E(y_{k-1} \mid n)$ in Equation (1) and solving the system, we obtain a reduced form equation that can be written,

$$y_k = \beta_{2,k-1} E(v \mid n) + \gamma_{2,k-1} v + \mu_n + \eta. \tag{3}$$

where η is proportional to u whereas the fixed effect μ_n is a linear combination of ϵ_n and u_n. Also, the reduced-form parameter $\beta_{2,k-1}$ is a linear combination of the $(\alpha_{k-t}\beta_{1,k-t} + \beta_{k-t})$ parameters whereas the $\gamma_{2,k-2}$ parameter is a linear combination of γ_{k-t} parameters, $t = 1, ..., k-6$ (see Technical Appendix).

The $\beta_{2,k-1}$ parameter captures the cumulative impact of neighbours' average maturity on a child's outcome at age k. Given that v and $E(v \mid n)$ are uncorrelated with u and ϵ_n, they are uncorrelated with the reduced-form residuals η and μ_n and Equation (3) shows that $\beta_{2,k}$ can be estimated through an OLS regression of y_k on $E(v \mid n)$. Intuitively, this reduced-form effect provides us with direct evidence of the existence of contextual effects. The $\beta_{2,k}$ parameter

[5] The initial condition of the recursive definition of y_k (k=7,...,16) is obtained by setting θ_6 and α_6 equal to zero

is indeed positive if and only if there is a t such that either α_{k-t} or β_{k-t} is positive.

Hence, observing the distribution of date of birth within small neighbourhoods makes it possible to identify contextual effects. Endogenous effects cannot be separated from exogenous contextual effects without an additional identifying assumption, however. One such additional assumption is that the date of birth within the year, as such, has no significant influence on own late school transitions and on that of neighbours. Using the notation of Equation (1), this amounts to assuming that β_{15} and γ_{15} are neglibible. Under this additional assumption, the distribution of neighbours' date of birth is clearly a valid instrumental variable for identifying the effect α_{15} of neighbours' early outcomes $E(y_{15} \mid n)$ on an adolescent's late outcome y_{16} whereas an adolescent's own date of birth is a valid instrument[6] for identifying the effect θ_{15} of own early outcome y_{15} on own late outcome y_{16}.

Assuming that respondents' dates of birth can be characterized by (say) two variables (v' and v'') rather than by just one (v), our identifying assumption can be tested as an overidentifying restriction. If the date of birth within the year has no effect on current outcome y_{16} on top of its effect on early outcome y_{15}, then any characterization v' of the date of birth is a valid instrument for identifying the effect of y_{15} on y_{16} and any further characterization v'' should have no additional effect on y_{16}. In what follows, we characterize date of birth by two dummies, "born between January and May" (v') and "born between June and November" (v'') (December being the reference). We show that the overidentifying restriction is not rejected. Also we check that the IV results remain unchanged when we characterize date of birth within the year by a single continuous variable rather than by one or two dummies (and the distribution of date of birth in the neighbourhood by its mean value rather by one or two proportions).

3.1 Results

Table 1 focuses on our basic sample of adolescents and analyzes their educational advancement at the end of junior high-school as a function of their own date of birth and of the date of birth of other adolescents living in the same neighbourhood. The first column shows that an adolescent's probability of being held back a grade at the end of junior high-school is about 8 percentage points larger (+16% of a SD) if the other adolescents living in the neighbourhood were born at the beginning rather than at the end of the year. Interestingly, the reduced-form effect of neighbours' maturity is almost as strong as that of own maturity.

[6] Notice that if θ_{15} were known, the identification of the endogenous effect β_{15} would only require an assumption that own date of birth, as such, does not affect neighbours' late transitions. To check the robustness of our results, it is possible to estimate β_{15} conditional on various value of θ_{15} (varying from 0 to 1) using neighbours' date of birth as the only instrumental variable. This strategy provides us with an upper and a lower bound for the parameter of interest β_{15} using a weaker exclusion restriction.

Having been held back a grade at the end of junior high-school is a cumulative outcome[7]. Hence, the effect estimated in column 1 represents the cumulative influence of close neighbours on both early and late educational outcomes. The second regression isolates the effect of close neighbours on late outcomes, as measured by the probability of grade repetition between the age of 15 and 16. It shows that an adolescent's probability of being in the same grade at age 16 as at age 15 is 5 percentage points larger (i.e. about 13% of a SD) if the other adolescents in the neighbourhood were born at the beginning rather than at the end of the year. As it turns out, the reduced-form effect of neighbours' maturity keeps on being strong at the end of junior high-school.

This reduced-form analysis does not separate endogenous from exogenous effects. As discussed above, it is possible to further explore this issue by assuming that date of birth within the year mostly affects early educational advancement. Table 2 shows the result of a regression of an adolescent's educational advancement at age 16 on own educational advancement at age 15 and on that of neighbours, using own date of birth and that of neighbors as instrumental variables. It reveals a significant endogenous effect (i.e. $\beta_{15} = .33$) which suggests that a one standard deviation increase in the proportion of neighbours held back a grade at age 15 increases ceteris paribus the probability of being held back a grade at age 16 by about 11 percentage points, i.e. 20% of a SD (Table 2, Column 3).

It is plausible that an adolescent's educational advancement directly affects an intermediate mechanism (e.g. own studying behaviour), and not neighbours' performance at school directly. Under this assumption, the endogenous effect estimated in this paper reflects the effect of neighbours' studying behaviour on own studying behaviour, the identifying assumption being that an adolescent's date of birth affects the studying behaviour of neighbours only insofar as it has affected own behaviour.

3.2 Overidentification and exogeneity tests

Standard overidentification tests do not reject the validity of our identifying assumptions[8]. We have checked that the results of the IV regression remain almost exactly the same when we use only the first dummy (i.e. being born between January and May) and the first proportion (i.e., proportion of neighbours born between January and May) as instrumental variables, the second dummy (i.e., being born between June and November) and the second proportion (i.e. proportion born between January and May) being used as additional control variables. These two additional control variables have no significant effect on the outcome under consideration (Table 2, Column 4). Also, the IV results are almost identical when we characterise the distribution of neighbours'

[7]According to a recent survey on education conducted by the French Statistical Office, about 20% of French individuals born in 1976-1985 have repeated a grade in primary school, 17% at the beginning of junior high-school and 14% at the end of junior high-school.

[8]For a presentation of the over-identification and endogeneity tests used in this paper, see for example Wooldridge (2002), pages 118 to 124.

date of birth within the year by its mean value rather than by two proportions (see Table A5 in the appendix). As discussed above, these different tests are consistent with the assumption that date of birth, as such, has no direct effect on the probability of repeating a grade between age 15 and 16 on top of its effect on the probability of repeating a grade before the age of 15[9]. The fact that the estimated endogenous effect remains the same regardless of whether we exclude one or two characterisations of the distribution of neighbours' date of birth can also be interpreted as meaning that the linear specification of the endogenous effect is not rejected (see Liebman et al. 2004 for a similar argument). We have also checked that the result remains unchanged when we add various family background indicators as control variables, which is consistent with our instruments being uncorrelated with family background.

A Hausman test rejects (at the 5% level) the assumption that the proportion of neighbours held back a grade is exogenous. Column 2 of Table 2 confirms that the OLS estimate of the proportion of neighbours held back a grade is significantly lower than the IV estimate. This is something of a puzzle, since typically, endogenous neighbourhood selection is likely to lead to upward bias in the OLS coefficient [10]. One potential explanation for the IV/OLS difference is that late grade repetition (i.e. between age 15 and 16) mainly affects relatively bad students living in good neighbourhoods on the one hand, and good students living in bad neighbourhoods on the other. Good students in good neighbourhoods do not repeat grades whereas bad students living in bad neighbourhoods have already been held back one or two grades early in their school career and are not likely to be held back further in subsequent periods. Hence, holding past educational advancement constant, there is a negative correlation between the quality of a neighbourhood and adolescents' propensities to repeat grades. Such a correlation can generate strong attenuation bias, i.e. the difference in late grade repetition across good and bad neighbourhoods may be very weak even though the true neighbourhood effect is very strong. Another potential source of attenuation bias arises from errors that affect our measure of respondents' outcomes and, consequently, our measure of the distribution of outcomes in the neighbourhood. Given that the variance of the errors in the measure of average outcomes decreases with the number of individuals, we should observe a smaller attenuation bias for a sample of respondents with more neighbours. To test these different interpretations, Table A6 in the Appendix compares OLS

[9] It should be emphasised that the correlation between own relative age and own late grade repetition is the combination of a potentially positive direct effect (relatively old children are more mature and perform better) and a potentially negative indirect effect (relatively old children are more likely to have already repeated a grade and, because of that, less exposed to further grade repetition). Hence, under the assumption that the direct effect is zero, the correlation between relative age and late grade repetition is not necessarily zero, but cannot be negative. Interestingly, the reduced-form effect of own date of birth in Table 2 column 1 shows that this correlation is actually positive, which means that our identifying assumption is not rejected.

[10] Interestingly, comparing experimental and non-experimental estimates, Liebman et al. (2004) do not find evidence of upward bias from non-random sorting of households across neighbourhoods, as would occur under the assumption that persons with good unobservables also have good outcomes and live in good neighbourhood.

estimates obtained using the full sample (column 1), with those obtained using the subsample of respondents with at least four neighbours (column 2) and those obtained on the subsample further restricted to the 15 year-old respondents who are still normal-age (i.e. neither ahead nor held back) and, consequently, the most exposed to late grade repetition (column 3). Interestingly, the OLS estimates are much larger in the more restricted sample (.24 in the third sample) and no longer different from the IV. This seems consistent with our interpretation of the IV/OLS gap.

3.3 Alternative specifications

Table 3 provides alternative evaluations of the endogenous effect using a non-cumulative specification of the dependent variable (column 1) and an alternative specification of the model (columns 2 to 3). Specifically, the first column shows the regression of an adolescent's probability of being in the same grade at age 15 and 16 (i.e. non-cumulative outcome) on the proportion of neighbours held back a grade at age 15 and own educational advancement at age 15, using the same instruments as in Table 2 (i.e. own date of birth within the year and that of neighbours). The estimated endogenous effect is as significant with this specification as with the cumulative one. A one SD increase in the proportion of neighbours held back a grade at age 15 increases an adolescent's probability of grade repetition between age 15 and age 16 by about 14 percentage points.

As discussed above, an alternative strategy is to estimate the endogenous effect β_{15} conditional on various values of θ_{15} in [0,1], using the distribution of neighbours' date of birth as the only instrument. If θ_{15} is assumed equal to 0 then β_{15} can be estimated through the IV regression of own educational advancement at age 16 on neighbours' educational advancement at age 15. In contrast, if θ_{15} is assumed equal to 1 then β_{15} can be estimated through the IV regression of grade repetition between age 15 and 16 on neighbours' educational advancement at age 15[11]. Table 3 shows these different regressions. They confirm that a one SD increase in the proportion of neighbours held back a grade at age 15 has a strong effect on own educational advancement at age 16. It lies between 22 percentage points when θ_{15} is assumed equal to 0 (column 3) and 10 percentage points when it is assumed equal to 1 (column 2). The estimated endogenous effect remains strong even under the extreme assumption that past educational advancement does not affect the current probability of grade repetition.

[11] If θ_{15} were assumed equal to (say) .5 then the dependent variable would be the mean of the two previous ones. We have checked that when we use this specification we obtain an evaluation of the endogenous effect which lies in between those obtained by setting $\theta_{15} = 0$ or $\theta_{15} = 1$.

4 Identification using information on families living in public housing

The previous section has focused on one specific contextual effect - the effect of performance at school of the other adolescents living in the neighbourhood. This section provides a broader evaluation of contextual effects, but without separating the endogenous from the exogenous dimensions. Specifically, we ask whether a child's performance at school is influenced by the level of human capital of families living in the neighbourhood, but will not address whether this is because it has a direct effect (exogenous channel) or because it affects the performance of other children in the neighbourhood (endogenous channel). This section uses available information on families living in public housing (HLM, about 20% of the population).

In France, any family is eligible for an HLM provided that the head of the family is allowed to live in France and that income per unit of consumption is below a threshold (about 30,000 Euros for a four person family in 2002) which depends on the region and which is updated at beginning of each year. Eligible families can apply for an HLM in any city (*commune*) where such public programs exist, regardless of their current place of residence or nationality.

Public housing is managed by several different types of administrative authority and - in general - eligible families apply simultaneously through the various possible channels. According to the Housing Survey conducted by the French Statistical Office in 2002, about 1.1 million households are waiting for public housing, whereas only about 400,000 such dwellings are made available each year. Hence, the waiting lists are very long and typically families have to wait for two or three years before a decision is made. Rents are considerably lower in public housing than in private-sector housing (-40% on average) which explains the high level of demand for public housing and the low level of turnover, especially in large cities (Laferrere and Leblanc, 2002). Within this framework, HLM managers have a very limited set of dwellings to offer each year to HLM applicants and very limited control over the neighbourhoods where the supply of dwellings is located. Families have even less control over the specific location of the dwelling to which they are allocated. Given these facts, the sorting of families across HLM neighbourhoods is plausibly much more exogenous than across private sector neighbourhoods.

To test this assumption, the first two columns of Table 4 focus on children who have just moved into a neighbourhood and show the results of a regression of a dummy indicating whether they have been held back a grade on the proportion of children who have been held back a grade in the neighbourhood into which they move. The first regression focuses on non-HLM neighbourhoods and reveals a very significant correlation between the two variables (column 1). Families who choose (or who are constrained by housing prices) to live close to one another are similar with respect to some important individual determinants of performance at school. Further explorations of the data (not reported) reveal that this correlation is mostly due to the fact that families who move into a non-

HLM neighbourhood and other families in this neighbourhood have similar levels of education and are likely to share the same nationality. When we add parental education and nationality as supplementary control variables, the effect found in the first column of Table 4 becomes very small and not significantly different from zero[12]. The second column shows the results of the same regression, but only for families moving into a HLM. Most interestingly, it shows that there is no correlation between the probability of being held back a grade for children moving into public housing and that for other children in the neighbourhood. The assignment of families across HLM neighbourhoods appears to be random with respect to children's educational performance.

In theory, the composition of HLM neighbourhoods could be biased by selective out-migration, even if the initial assignment were perfectly random. If this assumption were true, however, the correlation between the outcomes of children who have been living in the same HLM neighbourhood for more than one year would be driven (at least in part) by the similarity of their family background. Columns 4 and 6 of Table 4 focus on HLM families who have been living in their neighbourhood for more than one year and show that the correlation between children's performance at school and that of their neighbours does not decrease significantly when we control for their family background. This result does not hold true in non-HLM neighbourhoods, where we observe a very significant decrease in the regression coefficients when we control for the same set of family characteristics (Columns 3 and 5).

The rate of migration out of HLM housing is actually very low, because of the very low level of rents. According to the French Housing Survey, families observed in a HLM neighbourhood in 2002 had already spent an average of 10 years in their current residence, whereas the non-HLM families had spent only 5 years on average.

The findings reported in Table 4 are consistent with the existence of significant neighbourhood effects in HLM neighbourhoods and with the assumption that the HLM population is not sorted across neighbourhoods according to factors affecting early performance at school. Under this assumption, exogenous contextual effects can be evaluated in HLM neighbourhoods by standard OLS regressions. The first column of Table 5 focuses on HLM neighbourhoods and shows that an adolescent's advancement at school is negatively affected by the proportion of non-educated families in the neighbourhood. A one standard deviation increase in the proportion of non-educated neighbours generates a 6 percentage point increase in the probability of being held back a grade (12% of

[12] It should be emphasized, however, that the similarity of parents' education and nationality is not sufficient for explaining the correlation between the educational outcomes of adolescents who have been living in the same non-HLM neighborhood *for more than one year*. As shown below in column 5 of Table 5, the correlation between the performance of adolescents who have been living in the same neighborhood for more than one year remains significant and large even after controlling for the two main sources of endogenous neighborhood membership, i.e., parental education and nationality. A significant part of the observed correlation between the performance of children and the performance of their neighbors is due to endogenous neigborhood membership, but a significant part is not explained by this phenomenum and consistent with the existence of significant neighborhood effects.

a SD). In contrast, children's performance at school do not seem to be affected by the proportion of non-French families living in the neighbourhood[13].

The neighbourhoods with the highest proportions of non-educated families are also those with the highest proportions of single-parent families, the highest proportions of families with three or more children and also the highest proportions of unemployed adults. We have added these different neighbourhood characteristics as supplementary control variables in order to further explore the channels through which an adolescent's outcomes are influenced by the lack of education of other families in the neighbourhood. As it turns out, as shown by column 2 of Table 5, the proportion of single parent families and the proportion of large families have no effect, whereas the proportion of unemployed adults living in the neighbourhood has a significant effect. Column 3 adds the adolescent's early outcome as a supplementary control variable in order to separate the effect of context on current outcomes from the effect on early outcomes. The proportion of unemployed adults still has a significant effect whereas the effect of the proportion of non-educated families becomes non-significant. One interpretation of this set of results is that the proportion of non-educated parents affects adolescents' current outcomes mostly because it affects their early outcomes (plausibly through the endogenous channel) whereas the rate of unemployment in the neighbourhood also has a direct effect on current outcomes, maybe in part because it has a depressing effect on adolescents' incentives to pursue education.

5 Extension : the effect of early peers' dates of birth within the year

The French Labor Force Survey provides us with information on adolescents' outcomes only. To further explore the influence of social interactions on French children, we use a longitudinal survey recently conducted by the French Ministry of Education[14]. This survey provides detailed information on the early school career of a representative sample of pupils who started primary school in 1997. The basic sampling unit is the school. Within each school, a class in the first grade is drawn at random and a random sample of one third of new entrants are surveyed. Their performance is followed up until third grade. For about 7,500 pupils, we have information on their gender, exact date of birth and social background. We know their performance in the national tests that took place

[13] Oreopoulos (2003) examines the labor market outcomes of adults who were assigned to different public housing projects in Toronto (when children). He does not find a very significant long-run effect of having been assigned to relatively poor neighborhoods. In his paper, neighborhoods correspond to census tracts and contain about 1,000 to 3,000 housings. It is one potential explanation for the difference between his findings and ours. In the early Eighties, the French Statistical Office conducted a very interesting survey on interactions between neighbors which shows that French households interact on average with 2 or 3 very close neighbors only (see Héran, 1986)

[14] Piketty (2004) has used this dataset to explore the effect of class size on early school performance in France.

at entry into third grade and their performance in specific tests that took place in September 1997 at entry into first grade. Finally, we know the code of the 1997 school. Hence, for each respondent, we can identify the characteristics of his/her early classmates, i.e. pupils who were in the same class at entry into first grade. On average, we observe six early peers per respondent. All in all, this dataset makes it possible to analyze the relationship between performance in tests at entry into third grade and the date of birth of first-grade peers, using exactly the same specification as in the first section.

To begin with, we have checked that the observed individual characteristics of pupils are not correlated with the distribution of their peers' date of birth. In particular, there is no correlation between a child's sex, family background or date of birth within the year, on the one hand, and the proportion of early peers born at the beginning (or at the end) of the year on the other (see Table A7). This result confirms that pupils are not sorted in any systematic way across first-grade classes according to their date of birth within the year, i.e. the distribution of early peers' date of birth may be assumed exogenous to the respondent's own characteristics. Secondly, we have regressed pupils' scores in the tests conducted at entry into third grade on the distribution of their early peers' dates of birth, using the same specifications and control variables as in the LFS analysis (Table 6). The regressions have been performed for the global score and also separately for the scores obtained in mathematics and French. Most interestingly, this reduced-form analysis confirms that individual scores at entry into third grade decrease significantly with the proportion of first-grade peers who were born at the end of the year. A child's score at entry into third grade are 3 points (i.e. 20% of a SD=15 points) smaller when his/her early peers were born at the end of the year rather than during first months of the year. The effects are stronger and better estimated in mathematics than in French.

One potential problem with this analysis is that our measure of the proportion of peers born at the begining (or at the end) of the year is affected by sampling errors. The OLS estimates are affected by attenuation bias which increases with the variance of these errors. Given that early peers are randomly drawn among all the pupils in the class, the observed proportion of peers born at the beginning (or at the end) of the year is a consistent estimate of the true proportion, but its variance decreases with the number of peers actually observed in the survey n[15]. Hence, to evaluate the importance of the attenuation bias linked to sampling errors, we have replicated the previous analysis excluding the 15% of observations with the lowest n (i.e., with $n > 4$, see columns 4, 5, 6, Table 6). Comfortingly, the effects are stronger and better estimated for this subsample[16]. A child's score at entry into the third grade is about 26% of a SD smaller when his/her first-grade peers were born at the end of the year rather than during the first months of the year. The estimated effect remains larger in

[15] To be more specific, the variance of the errors is proportional to $\theta = \frac{(N-n)}{(N-1)n} \simeq \frac{1-t}{n}$, where $t = \frac{n}{N} = 1/3$ is the sampling rate used within classes.

[16] We have checked that the estimated effects do not increase when we further restrict the sample.

mathematics (33% of a SD) than in French (20%).

Test scores at entry into third grade represent a measure of the quality of the first two years at school. Assuming that a child's performance at school is affected by the date of birth within the year of his/her early peers only insofar as their date of birth within the year affects the quality of their own early schooling, we can use the distribution of peers' date of birth as an instrument to identify the true effect of peers' early school performance on a child's performance. Table 7 shows the OLS and IV regressions of a child's score at entry into third grade on the average score at entry into third grade of the 1^{st} grade peers. The IV estimates are significant and large: a one SD increase in the average score of early peers increases a pupil's score by about 36% of a SD. Overidentification tests do not reject our identifying assumption. The IV estimate is not significantly different from the OLS estimate, however.

The first sections of this paper show that an adolescent's outcomes at the end of junior high-school are strongly affected by the educational advancement of the other adolescents living in the same neighbourhood. This reflects interactions that mostly take place outside the classroom, since it is unlikely that two adolescents attend the same class within the same school, even when they are close neighbours[17]. The data from the Ministry of Education suggest that early interactions within French primary schools have no less influence on a child's educational career than interactions between adolescents in the same neighbourhood. The influence of close neighbours on own educational outcomes seems significant from the beginning to end of compulsory education, both inside and outside of the classroom'.

6 Conclusion

Buiding on the specificities of French institutions, we analyse the influence of close neighbours' characteristics on an adolescent's performance at school. Our first strategy builds on the fact that the date of birth within the year, as such, has a significant effect on early educational outcomes. We use the distribution of close neighbours' month of birth as an instrumental variable to identify the influence of neighbour's early outcomes on an adolescent's educational advancement at the end of junior high-school. This approach suggests that the probability of repeating a grade at the end of junior high-school increases strongly when the other adolescents living in the same neighbourhood have already been held back a grade rather than when they have not. A second strategy uses the

[17] To begin with, there is some flexibility in the choice of the junior-high school (called *collège).* As a consequence, two neighboring adolescents do not necessarily attend the same *collège* : about 20% of adolescents attend private *collège* and another 20% do not attend the nearest public *collège*. Given this fact, one can estimate that only about one third of neighboring adolescents actually attend the same *collège*. And even if they do, the probability remains weak that they will be found in the same class. According to the ministry of education, a typical *collège* has on average 10 classes for the eighth and ninth grades. Hence, the probability of finding in the same class two 15-years-olds attending the same *collège* is not more than about 10%.

fact that the distribution of families across public housing (HLM) is not significantly different from quasi-random assignment. In such a context, the influence of close neighbours' families can plausibly be identified through standard regressions. This strategy shows that an adolescent's educational advancement is negatively influenced by the proportion of non-educated families living in the neighbourhood.

Our paper focuses on the influence of close neighbours on performance at school. Further research is needed, however, to explore the effects of close neighbours on other outcomes, such as the decision to drop out of school or the decision to participate in the labor market. We speculate that neighbourhood effects for such decisions are even stronger than the neighbourhood effects on school performance *stricto censu.* Put differently, we speculate that close neighbours have more influence on own preferences than on own resources. Also further research is needed to better explore the channels through which children living in the same neighbourhood influence each other. It is obviously a key issue for defining public policies. In particular, it would be useful to better identify the contribution of social interaction during extra-curricular activities. Generally speaking, similar evaluations need to be performed in other countries to explore whether (and why) the role of social interaction varies across societies.

References

[1] Aaronson, D. 1998, "Using Sibling Data to Estimate the Impact of Neighborhoods on Children's Educational Outcomes", Journal of Human Resources, 23, 915-946.

[2] Allen, Jeremiah and Roger Barnsley, 1993, "Streams and Tiers: the Interaction of Ability, Maturity and Training in System with Age-dependent Recursive Selection", The Journal of Human Resources, 28(3), 649-659.

[3] Angrist, Joshua D. and Kevin Lang, 2004. "Does School Integration Generate Peer Effects? Evidence from Boston's Metco Program," American Economic Review, vol. 94(5), 1613-1634.

[4] Barnsley Roger H., Thompson Alan H. and Paula E. Barnsley, 1985, "Hockey Sucess and Birth Date: The Relative Age Effect", Canadian Association for Health, Physical Education, and Recreation Journal 51, 23–8.

[5] Brock, William A. and Steven N. Durlauf, 2001, "Interaction-Based Models", in Handbook of Econometrics vol:V, eds. James J. Heckman and Edward E. Leamer, Amsterdam: Noth-Holland, 3297-3380.

[6] Case, Anne C. and Lawrence F. Katz, 1991. "The Company You Keep: The Effects of Family and Neighborhood on Disad- vantaged Youths," NBER Working Papers 3705, National Bureau of Economic Research, Cambridge, MA.

[7] Cutler David M. and Edward L. Glaeser, 1997, "Are Ghettos Good or Bad? ", Quarterly Journal of Economics, 112 (2), 827-872.

[8] Ding, Weili and Steven F. Lehrer, 2004, "Do Peers Effect Student achievement in China's Secondary School", WP n°35, School of Policy Studies,Queens University,October

[9] Evans W. N., W. E. Oates et R. Schwab, 1992, "Measuring Peer Group Effects: A Study of Teenage Behavior ", Journal of Political Economy, 1000, 966-991.

[10] Gibbons, Stephen and Shqiponja Telhaj, 2005, Peer Effect and Pupil Attainment: Evidence from Secondary School Transitions, Unpublished Manuscript.

[11] Ginther, Donna, Robert Haveman, and Barbara Wolfe, 2000, "Neighborhood Attributes as Determinants of Children's Outcomes: How Robust are the Relationships?", Journal of Human Resources, 35(4), 603-642.

[12] Gould, Eric D., Lavy, Victor and M. Daniele, Paserman, 2004, "Immigrating to Opportunity: Estimating the Effect of School Quality using a Natural Experiment on Ethiopans in Israel", Quarterly Journal of Economics, 119 (2),489-526.

[13] Hanushek, Eric, John Kain, Jacob Markman and Steve Rivkin, 2003, "Does Peer Ability Affect Student Achievement?" Journal of Applied Econometrics, 18 (5), 527-544.

[14] Héran, François, 1986," Comment les Français voisinent ", Économie et statistique, n° 195, pp. 43-59.

[15] Hoxby, Caroline, M. (2001), "Peer Effects in the Classroom: Learning from Gender and race Variation", WP 7867, Cambridge, MA:NBER.

[16] Kremer, Michael and Levy, Dan, 2003, Peer Effects and Alcohol Use Among College Students, WP 9876, Cambridge, MA: NBER.

[17] Ioannides, Yannis M., 2002, "Residential Neighborhood Effects", Regional Science and Urban Economics, 32,2, 2002, 145-165.

[18] Ioannides, Yannis M., 2003, "Interactive Property Valuations", Journal of Urban Economics, 56, 435-457.

[19] Ionnadies Yannis M. and Jeffrey E. Zabel, 2003, "Neighborhood Effects and Housing Demand", Journal of Applied Econometrics, 18, 563-584.

[20] Jacob, Brian A., 2004, "Public Housing, Housing Voucher, and Student Achievement: Evidence from Public Housing Demolitions in Chicago", American Economic Review 94 (1), 233-258.

[21] Leventhal, Tama and Jeanne Brooks-Gunn, 2004, "A Randomized Study of Neighborhood Effects on Low-Income Children's Educational Outcomes," Developmental Psychology, 40 (4):488-507.

[22] Liebman, Jeffrey, Lawrence R. Katz, Jeffrey R. Kling, 2004, "Beyond Treatment Effects: Estimating the Relationship between Neighborhood Poverty and INdividual Outcomes in the MTO Experiment", WP 493, Industrial Relations Section, Princeton University.

[23] Ludwig Jens, Duncan Greg J. and Pinkson J.C, 2000, "Neighborhoods Effects on Economic Self-sufficiency: Evidence from a Randomized Housing-Mobility Experiment", Unpublished manuscript.

[24] Ludwig, Jens, Helen F. Ladd, and Greg J. Duncan, 2001, "Urban Poverty and Educational Outcomes", Brooking-Wharton Papers on Urban Affairs, 147-201.

[25] McEwan, Patrick, 2003, "Peer Effects on Student Achievement: Evidence from Chile", Economics of Education Review, 22 (2), 131-141.

[26] Manski Charles F., 1993, " Identification of Endogenous Social Effects : The Reflection Problem ", Review of Economic Studies, 60, 531-542.

[27] Moffitt, Robert A., 2001, "Policy Interventions, Low-level Equilibria and Social Interactions", in Social Dynamics,eds., Steven N. Durlauf and H. Peyton Young, Cambridge: MIT Press, 45-82.

[28] Murat F. and T. Rocher, 2003, "La Place du Projet Professionnel dans les Inégalités de Réussite Scolaire à 15 ans", in *France Portrait Social,* 2002-2003, INSEE, Paris,101-121.

[29] Oreopoulos, Philip, 2003, "The Long-Run Consequence of Living in a Poor Neighborhood", Quarterly Journal of Economics, 118 (4), 1533-1575.

[30] Piketty, Thomas, 2004, "L'impact de la Taille des Classes et de la Ségrégation Sociale sur la Réussite Scolaire dans les Ecoles Françaises : une Estimation à Partir du Panel Primaire 1997", Unpublished Manuscript.

[31] Pischke, Jörn-Steffen and Andreas, Ammermueller, 2006, Peer Effects in European Primary Schools: Evidence from PIRLS, Unpublished Manuscript.

[32] Rosembaum, James E., 1995, "Changing Geography of Opportunity by Expanding Residential Choice: Lessons from Gautreaux Program", *Housing Policy Debate*, 6 (1) : 231-269.

[33] Sacerdoce, Bruce, 2001, "Peer Effect With Random Assignment: Results for Darthmouth Roomates", Quarterly Journal of Economics, 116(2), 681-704.

[34] Sanbonmatsu, Lisa, Jeffrey R. Kling, Greg J. Duncan, and Jeanne Brooks-Gunn, 2004, "Experimental Analysis of Neighborhood Effects", Princeton Industrial Relations Section Working Paper 492, August.

[35] Wooldridge, Jeffrey M., 2002, *Econometrics Analysis of Cross Section and Panel Data*, MIT Press.

[36] Zimmerman, David J., 2003, "Peer Effects in Academic Outcomes: Evidence from a Natural Experiment", Review of Economics and Statistics, 85(1), 9-23.

Technical Appendix

This appendix expresses the parameters of Equations (2) and (3) as a function of the parameters of Equation (1). To begin with, after averaging, Equation (1) yields,

(A1) $E(y_k \mid n) = (\alpha_{k-1}+\theta_{k-1})E(y_{k-1} \mid n)+(\beta_{k-1}+\gamma_{k-1})E(v \mid n)+\epsilon_n+u_n$,

which can be rewritten,

(A2) $E(y_k \mid n) = \beta_{1,k-1}E(v \mid n) + \gamma_{1,k-1}(\epsilon_n + u_n)$,

where,

$\beta_{1,k-1} = (\beta_{k-1}+\gamma_{k-1}) + \sum_{2\leq t\leq k-6}[(\alpha_{k-1}+\theta_{k-1})...(\alpha_{k-t+1}+\theta_{k-t+1})(\beta_{k-t}+\gamma_{k-t})]$ and

$\gamma_{1,k-1} = 1 + \sum_{2\leq t\leq k-6}[(\alpha_{k-1}+\beta_{k-1})...(\alpha_{k-t+1}+\theta_{k-t+1})]$.

Using (A2), Equation (1) implies,

(A3) $y_k = (\alpha_{k-1}\beta_{1,k-1}+\beta_{k-1})E(v \mid n)+\gamma_{k-1}v+\theta_{k-1}y_{k-1}+(\alpha_{k-1}\gamma_{1,k-1}+1)\epsilon_n+\alpha_{k-1}\gamma_{1,k-1}u_n+u$,

which can be rewritten,

(A4) $y_k = \beta_{2,k-1}E(v \mid n) + \gamma_{2,k-1}v + \delta_{2,k-1}\epsilon_n + \lambda_{2,k-1}u_n + \tau_{2,k-1}u$,

where,

$\beta_{2,k-1} = (\alpha_{k-1}\beta_{1,k-1}+\beta_{k-1}) + \sum_{2\leq t\leq k-6}[\theta_{k-1}...\theta_{k-t+1}(\alpha_{k-t}\beta_{1,k-t}+\beta_{k-t}))]$,

$\gamma_{2,k-1} = \gamma_{k-1} + \sum_{2\leq t\leq k-6}[\gamma_{k-t}\theta_{k-1}...\theta_{k-t+1})]$,

$\delta_{2,k-1} = (\alpha_{k-1}\gamma_{1,k-1}+1) + \sum_{2\leq t\leq k-6}[\theta_{k-1}...\theta_{k-t+1}(\alpha_{k-t}\beta_{1,k-t}+\beta_{k-t}))]$,

$\lambda_{2,k-1} = \alpha_{k-1}\gamma_{1,k-1} + \sum_{2\leq t\leq k-6}[\theta_{k-1}...\theta_{k-t+1}(\alpha_{k-t}\gamma_{1,k-t}))]$,

$\tau_{2,k-1} = 1 + \sum_{2\leq t\leq k-6}[\theta_{k-1}...\theta_{k-t+1}]$.

Table 1: Reduced-form effect of close neighbours' date of birth within the year on an adolescent's educational advancement.

Independent variables	Dependent variable :	
	[Held Back a Grade at 16=1]	[Grade Repetition between 15 and 16=1]
Characteristics of the other 15-years-old living in the aire.		
Proportion born January-May	-.082 (.027)	-.050 (.022)
Proportion born June-November	-.050 (.027)	-.032 (.022)
Proportion born December	Ref.	Ref.
Individual characteristics		
Born January-May	-.085 (.016)	.032 (.014)
Born June-November	-.049 (.016)	.020 (.013)
Born December	ref.	ref.
Additional controls (1)	yes	yes
R^2	.04	.05
Nb. Obs.	13,116	13,116

Source : LFS, t=1991 to 2002, Insee.
Sample : respondents born in *t-15* , observed at *t* and *t+1*, who have been living in their neighbourhood for more than one year. Standard deviation in brackets.
(1) Additional controls include year, gender and nationality dummies.

Table 2: The effect of close neighbours' educational advancement on an adolescent's educational advancement: an evaluation using the distribution of close neighbours' date of birth as an instrumental variable.

Independent variables	Proportion of other adolescents held back a grade at 15	OLS	IV	IV
		Dependent variables : [Held back a grade at age 16=1]		
Characteristics of the other 15-years-old living in the aire				
Prop. held back a grade at 15	-	.08 (0.01)	.33 (.13)	.36 (.15)
Prop. born January-May	-.14 (0.02)	-	-	-
Prop. born June-November	-.07 (0.02)	-	-	-.006 (.012)
Prop. born December	Ref.	-	-	Ref
Individual characteristics				
[Held back a grade at 15=1]	-	.70 (.01)	.57 (.08)	.57 (.08)
Born January-May	-.009 (.010)	-	-	-
Born June-November	-.002 (.010)	-	-	-.005 (.007)
Born December	Ref.	-	-	Ref.
Additional Controls (1)	yes	yes	yes	yes
R^2	.03	.06	.06	.06
H_0 = "Proportion held back a grade at 15" exogenous"	-	-	-.25 (.13) (reject..5%)	-.28 (.15) (reject. 6%)
H_0 = Instruments jointly valid	-	-	.22 (.80) (not reject.)	.00 (1.00) (not reject)
Nb. Obs.	13,116	13,116	13,116	13,116

Source : LFS, t=1991 to 2002, Insee. Sample : respondents born in *t-15* , observed at *t* and *t+1*, who have been living in their neighbourhood for more than one year. Standard deviation in brackets.
(1) Additional controls include year, gender and nationality dummies.
Note: The two potentially endogenous regressors are the proportion of other adolescents held back a grade at age 15 and a dummy indicating whether the individual is held back a grade at age 15. The instruments are two proportions characterizing the distribution of neighbours' dates of birth within the year and two dummies characterizing individual date of birth within the year. The first column shows the (first-stage) regression of the proportion of neighbours held back a grade on the instruments. The last column shows an IV regression where the proportion of neighbours born between June-November and the dummy indicating whether the respondent was born between June-November are used as additional control variables rather than as instrumental variables.

Table 3: The effect of close neighbours' educational advancement on an adolescent's educational advancement: alternative dependent variables and alternative specifications.

	[Grade repetition Between 15 and 16=1]	[Grade repetition Between 15 and 16=1]	[Held back a grade at 16=1]
Proportion other adolescents in the *aire* held back a grade at age 15	.43 (.15)	.33 (.14)	.72 (.20)
[Held back a grade at age 15=1]	-.24 (.09)	-	-
Born January-May	-	.035 (.014)	-.10 (.02)
Born June-November	-	.021 (.013)	-.05 (.02)
Additional Controls (1)	yes	yes	yes
R^2	.01	.01	.04
Number of Observations	13,116	13,116	13,116

Source : LFS, t=1991 to 2002, Insee. Sample : respondents born in *t-15* , observed at *t* and *t+1*, who have been living in their neighbourhood for more than one year. Standard deviation in brackets.

(1) Additional controls include year, gender and nationality dummies.

Note: Column 1 shows an IV regression where the dependent variable is a dummy indicating whether an adolescent is in the same grade at age 15 and 16 and where the potentially endogenous regressors (i.e., neighbours' and own educational advancement at 15) and the instruments (neighbours' and own date of birth) are the same as in Table 2. Columns 2 and 3 replicates previous IV analysis with the effect of own educational advancement at age 15 being set to zero and dummies indicating own date of birth used as additional controls variable rather than as instruments.

Table 4: Endogenous neighbourhood membership in HLM and non-HLM neighbourhoods.

	Dependent Variable : [Held back a grade at 15=1]					
	Sample who just moved into the neighbourhood		Sample who have been living in the neighbourhood for more than one year			
	Private (non-HLM)	Public (HLM)	Private (non-HLM)	Public (HLM)	Private (non-HLM)	Public (HLM)
Proportion other adolescents in the *aire* held back a grade at 15	15 (.04)	-.02 (.10)	.20 (.01)	.20 (03)	13 (.01)	18 (.03)
Additional Controls (1)	no	no	no	no	yes	yes
R^2	.03	.07	.05	.04	.12	.06
Nb Obs.	1,096	245	10,759	2,356	10,759	2,356

Source : LFS, t=1991 to 2002, Insee. Samples : In columns 3 to 5, respondents born in *t-15*, who have been living in their neighbourhood for more than one year. In columns 1 and 2, respondents born in *t-15*, who have just moved into their neighbourhood.

(1) All regressions include years and gender dummies. Additional controls in columns 5 and 6 include a nationality dummy and four father's education dummies. Standard deviation in brackets.

Table 5 : Exogenous contextual effects: a reduced-form evaluation using information on the families who live in public housing.

	Dependent variable : [Held back a grade at 16=1]		
Characteristics of the other families living in the aire			
Proportion high-school dropouts	.15 (.04)	.12 (.05)	.05 (.04)
Proportion non-French	-.05 (.05)	-.06 (.05)	-.04 (.04)
Proportion unemployed	-	.21 (.07)	.15 (.06)
Proportion single-parent family	-	-.07 (.09)	-.06 (.07)
Proportion large families (3 or more children)	-	-02 (.06)	-02 (.05)
Individual characteristics			
[Held back a grade at 15=1]	-	-	.41 (.13)
Additional controls (1)	yes	yes	yes
R^2	.07	.08	.13
Nb Obs.	2 356	2 356	2 356

Source : LFS, 1991 to 2002, Insee. Sample : respondents born in *t-15* living in a HLM neighbourhood and who have been living in the neighbourhood for more than one year. Standard deviation in brackets.

(1) Additional controls include year, nationality, gender, father's unemployment, father's and mother's education dummies.

Table 6 : The effect of the distribution of 1st grade peers' dates of birth on a pupil's performance at the entry into the 3rd grade.

	Dependent Variable: Score at entry into the 3rd Grade					
	Total Sample			Sample excluding 15% largest sampling errors (i.e., number *1st* grade peers>4)		
	Total score	Score in French	Score in Math	Total score	Score in French	Score in Math
Date of birth of 1st Grade peers						
Prop. Jan.-May	3.0	2.0	4.1	3.9	3.1	4.9
	(1.1)	(1.2)	(1.2)	(1.3)	(1.4)	(1.4)
Prop. June-Nov.	1.9	0.7	3.2	2.2	1.2	2.2
	(1.1)	(1.2)	(1.2)	(1.3)	(0.4)	(1.2)
Prop. Dec.	Ref.	Ref.	Ref.	Ref.	Ref.	Ref.
Individual characteristics						
Boy	-1.4	-3.7	1.0	-1.3	-3.7	1.0
	(0.3)	(0.3)	(0.3)	(0.3)	(0.3)	(0.3)
Additional controls (1)	yes	yes	yes	yes	yes	yes
R^2	.46	.41	.41	.46	.41	.41
Nb. Obs.	7450	7512	7501	6262	6314	6303

Source: Panel Primaire 1997, French Ministry of Education. Sample : French pupils who enter into primary school in September 1997. Standard deviation in brackets.
(1) Additional controls include the first-grade score, the country of birth, six dummies indicating father's socioeconomic status and dummies indicating whether the individual was born in Jan-May, June-Nov. or Dec.

Table 7 : The endogenous effect on test scores at entry into 3rd grade: an evaluation using the distribution of dates of birth of 1st grade peers as an instrumental variable.

	First-Stage	Dependent Variable: Individual score at entry 3rd grade	
		OLS	IV
Average Score of 1st grade peers at entry into 3rd grade	-	0.27 (0.01)	0.36 (0.11)
Dates of birth of 1st *grade peers:*			
Prop. January-May	6.9 (0.8)	-	-
Prop. June-November	2.8 (0.8)	-	-
Prop. December	Ref.	-	-
Additional controls	Yes	yes	yes
H_0 = Peers' average score exogenous		-	-.23 (.11) (reject..5%)
H_0 = Instruments Valid		-	.18 (.83) (not reject.)
Nb Obs.	7279	7279	7279

Source: Panel Primaire 1997, French Ministry of Education. Sample : French pupils who enter into primary school in September 1997. Standard deviation in brackets.
(1) Additional individual controls : first-grade score, country of birth, six dummies indicating father's socioeconomic status and dummies indicating whether the pupil was born in Jan-May, June-Nov. or December.
Note: The first column shows the (first stage) regression of first-grade peers' average score on the distribution of their dates of birth. Column 3 shows the IV regression of a pupil's score at entry into third grade on peers' average score using the distribution of peers' dates of birth as instrumental variable. Column 2 shows the corresponding OLS.

Appendix A

Table A1 : Distribution of LFS 15-years-old respondents according the number of other 15-years-old children living in their *aire*.

Nb of other 15-years-old adolescents in the *aire*	Nb of 15-years-old respondents	Proportion in the population of 15-years-old respondents
1	1731	13.2
2	2270	17.3
3	2076	15.8
4	1793	13.7
5	1293	9.9
6	990	7.6
7	777	5.9
8	594	4.5
9	453	3.4
10 or more	1139	8.7
Total	*13116*	*100*

Source: LFS *t*=1991,…2002. Sample: 15-years-old respondents, observed at *t* and *t+1*, who have been living in their neighbourhood for more than one year. Standard deviation in brackets.

Table A2 : Descriptive statistics.

	Mean	Standard-deviation
Individual Characteristics		
Held back a grade at 15	.42	.49
Held back a grade at 16	.57	.50
Same grade at 15 and 16	.20	.40
Boy	.51	.50
Non-french	.06	.23
Born January-May	.42	.49
Born June-November	.50	.50
Father high-school dropout	.32	.46
Characteristics of the other 15-years-old respondents living in the aire		
Proportion held back a grade	.42	.32
Proportion born January-May	.42	.30
Proportion born June-November	.49	.30
Proportion parents high-school dropout	.38	.27
Proportion Non-French	.06	.11
Number of Observations	*13116*	

Source: LFS *t*=1991,…2002.
Sample: 15-years-old respondents observed at *t* and *t+1*, who have been living in their neighbourhood for more than one year. Standard deviation in brackets.

Table A3: Relationships between an adolescent's characteristics and the distribution of dates of birth of other adolescents in the neighbourhood.

Independent variables	Dependent variables :		
	Proportion neighbours born January-May	Proportion neighbours born June-December	Neighbours' average month of Birth
Date of birth (continuous specification	-	-	.005 (.005)
Date of birth (dummies)			
Born Janurary-May	.008 (.010)	-.006 (.010)	-
Born June-November	.001 (.010)	.001 (.010)	-
December	Ref.	Ref.	-
Boy	-.008 (.005)	.013 (.05)	.001 (.036)
Non-French	.002 (.012)	.001 (.012)	-.007 (.008)
Father's education			
College grad.	-.004 (.009)	.005 (.009)	.009 (.063)
High-school grad.	.010 (.010)	-.011 (.011)	-.051 (.072)
Vocational	Ref.	Ref.	Ref.
No Dip.	-.010 (.009)	.008 (.009)	.022 (.063)
Missing	-.003 (.010)	.007 (.010)	.028 (.070)
Mother's education			
College grad.	-.000 (.009	.004 (.009)	.009 (.063)
High-school grad.	.006 (.009)	-.003 (.010)	-.025 (.066)
Vocational	Ref.	Ref.	Ref.
No Dip.	-.002 (.009)	-.006 (.009)	.050 (.059)
Missing	-.002 (.009)	.016 (.009)	.077 (.063)
Nb. Obs.	13,116	13,116	13,116
R^2	.002	.003	.001
Fisher (4 dummies father Educ.=0)	.93 (.42)	.83 (.47)	.46 (.76)
Fisher (4 dummies mother Educ.=0)	.17 (.91)	.28 (.83)	.42 (.74)

Source: LFS t=1991,…2002. Sample: 15-years-old respondents, observed at t and t+1, who have been living in their neighbourhood for more than one year. All regressions include a set of eleven years dummies as additional control variables. Standard deviation in brackets.

Table A4: Adolescents' characteristics and neighbours' dates of birth.

Individual characteristics	Distribution of dates of birth of 15-years-old neighbours	
	Prop. born January-May	Prop. born June-November
Father college grad.	42.4 (.7)	49.7 (.7)
Father not college grad.	42.3 (.3)	49.6 (.3)
Born Jan-May	42.7 (.4)	49.2 (.4)
Born June-Nov.	42.0 (.4)	49.9 (.4)
Boy	42.3 (.6)	49.0 (.6)
Girl	42.2 (.6)	50.6 (.6)
French	42.3 (.2)	49.6 (.2)
non French	41.7 (1.1)	50.9 (1.1)

Source: LFS t=1991,…2002. Sample: 15-years-old respondents, observed at *t* and *t+1*, who have been living in their neighbourhood for more than one year. Note: The average proportion of peers born in January-May is 42.3% for boys and 42.2% for girls.

Table A5: The endogenous contextual effect: an evaluation using average month of birth as an instrumental variable.

Independent variables		Dependent var. : [Held back a grade at 16=1]		
	First stage (1)	Reduced form	OLS	IV
Characteristics of the other 15-years-old living in the aire				
Prop. held back a grade at 15	-	-	.08 (.01)	.31 (.12)
Individual Characteristics				
[Held back a grade at 15=1]	-	-	.70(.01)	.61(.07)
Average month of birth (continuous specification)	.012 (.01)	-.006 (.0020)	-	-
Month of Birth (continuous specification)	.0002 (.0008)	-.008 (.001)	-	-
Cohort, Gender, nation. dummies	yes	yes	yes	Yes
R^2	.03	.04	.51	.06
Number Observations	13,116	13,116	13,116	13,116

Source: LFS t=1991,…2002. Sample: 15-years-old respondents, observed at t and t+1, who have been living in their neighbourhood for more than one year. Note: This table shows an analysis where the dependent variable and the potentially endogenous regressors are the same as in Table 2, but where the instruments are neighbours' average month of birth and own month of birth (m=1,2…12). The first column shows the (first stage) regression of the proportion of neighbours held back a grade on their average month of birth and individual characteristics.

Table A6 : Variation in OLS estimates of the endogenous effect across sub-samples.

	Dependent variable: [Held back a grade at 16=1]		
	Full Sample	Sub-sample A	Sub-sample B
Prop. Held back at 15	.08 (01)	.13 (.02)	.25 (.03)
Held back at 15	.70 (.01)	.69 (.01)	-
Nb. Obs.	13116	5246	2915

Source: LFS t=1991,…2002. Full sample: 15-years-old respondents, observed at t and t+1, who have been living in their neighbourhood for more than one year. Standard deviation in brackets. Regressions include a set of eleven years dummies as additional control variables. Standard deviation in brackets.
Subsample A : respondents with 5 or more other adolescents in the neighbourhood. Subsample B: respondents who are normal-age at t-1 and with 5 or more other adolescents in the neighbourhood.
All regressions include a set of eleven years dummies, a gender dummy, a nationality dummy and two date-of-birth dummies as additional control variables.

Table A7: Pupils' characteristics and 1st grade peers' dates of birth.

Pupils' characteristics	Distribution of 1st Grade Peers' Months of Birth	
	Prop. born January-May	Prop. born June-November
Father college grad.	41.7 (.5)	49.9 (.5)
Father not college grad.	41.2 (.2)	50.5 (.5)
Born January-May	41.5 (.3)	50.1 (.3)
Born June-November	41.1 (.3)	50.6 (.3)
Boy	41.2 (.3)	50.6 (.3)
Girl	41.3 (.3)	50.2 (.3)
French	41.2 (.2)	50.4 (.2)
Non French	43.7 (1.5)	48.8 (1.5)

Source: Panel Primaire 1997. Sample : French pupils who enter into primary school in September 1997. Standard deviation in brackets.
Note: The average proportion of peers born in January-May is 41.2% for boys and 41.3% for girls.