Revista de Economía Aplicada
ISSN: 1133-455X
rea@unizar.es
Universidad de Zaragoza
España

GUTIÉRREZ-DOMÈNECH, MARIA; ADSERÀ, ALÍCIA
STUDENT PERFORMANCE IN ELEMENTARY SCHOOLS
Revista de Economía Aplicada, vol. XX, núm. 59, 2012, pp. 135-164
Universidad de Zaragoza
Zaragoza, España

Revista de Economía Aplicada E A *Número 59 (vol. XX), 2012, págs. 135 a 164*

STUDENT PERFORMANCE IN ELEMENTARY SCHOOLS*

MARIA GUTIÉRREZ-DOMÈNECH
La Caixa

ALÍCIA ADSERÀ
Princeton University

This paper studies the association between children's cognitive and non-cognitive development with socioeconomic factors, and school characteristics in Catalonia. The focus of the analysis is on the relevance of the quarter of birth and the school entrance age. The data, collected in 2005, covers children attending 2nd, 4th and 6th grade in a random sample of 191 schools. We find that children born late in the year, close to the December 31st cutoff date, display lower academic performance than those born in the first two quarters, but no difference in social behavior appears by quarter of birth. The maturity gap does not seem to disappear as children advance into later grades. Children who attended nursery school generally do better than those who started school either at preschool (around 3 years old) or later. This relationship weakens when origin and family education controls are included in the analysis. Among other demographic characteristics, children raised in non-nuclear families, with low educated parents, with little exposure to after school activities or born outside Spain tend to underperform others at school.

Key words: educational economics, quarter of birth, school performance, family, use of time, entrance age.

JEL classification: I20, J24.

This paper analyses the association between children's cognitive and non-cognitive development in Catalonia with socioeconomic factors and school characteristics, with a focus on quarter of birth and entrance age. Understanding what matters for academic achievement is a prerequisite to designing best practices to improve the educational attainment of the population.

(*) We would like to thank "La Fundació Jaume Bofill" for providing the data and to participants in seminars at Universitat Pompeu Fabra, INSIDE conference at IAE (Bellaterra), and at the European Workshop on Labour Markets and Demographic Change in the Vienna Institute of Demography for their comments. The views expressed in this study are those of the authors only and do not necessarily represent those of "la Caixa". Correspondence to: Maria Gutiérrez-Domènech, "la Caixa" -Research Department (9012), Av. Diagonal 629, torre I, 08028 Barcelona. E-mail: maria.gutierrez.domenech@lacaixa.com. Fax: +34934046892.

The growing skill premium during recent years is, in part, an outcome of the increasing importance of education in today's labour market. The emergence of new technologies that drive economic growth demands, in addition to ideas and invention, a qualified workforce that is able to use modern resources. The relationship between human capital and growth is well established [Nelson and Phelps (1966), Romer (1990) and Rebelo (1991)]. More recently, research has focused on the association between labour-force quality and growth. Hanushek and Kimko (2000), for instance, show that direct measures of labour-force quality from cross-country comparative mathematics and science tests are strongly related to economic growth. Consequently, improving the cognitive and non-cognitive skills of the population constitutes a fundamental policy for governments.

This paper uses data collected in 2005 for the project "Família i Educació a Catalunya" by the Fundació Jaume Bofill. Cognitive development is measured with two quantitative indicators reported by the teacher: *Global knowledge* and *Catalan Knowledge* (the vehicular language of instruction in elementary schools in Catalonia). Noncognitive knowledge is measured with two indicators: *School Abilities* (reported by the teacher) and *Social Behaviour* (a combination of several qualitative measures given by both teachers and parents). Children of three different grades of elementary school (second, fourth and sixth grade) were surveyed. In addition to quarter of birth and age at which the child first attended school, the models estimated also include controls for geographic origin, language spoken at home, income level, educational attainment and labour force status of the parents as well as their values and political attitudes, family composition, school characteristics and time allocated to after-school activities.

The main focus of this paper is on understanding whether relative age within a classroom –measured here by quarter of birth– and the age at which a child had his first educational experience outside the home (in any form of daycare, nursery or formal schooling) are associated with differential cognitive and non-cognitive outcomes. There is a large literature in education and economics interested on those issues. Many papers find differentially low academic outcomes among children born just before the cutoff age and who are, as a result, the youngest in their classrooms. Cascio and Schanzenbach (2007), in a US study, find that this is only valid for disadvantaged students who, when placed among older children in the same classroom, are less likely to take a college-entrance exam than other kids of their exact same age.

Further, the findings are not conclusive as to whether those initial differences persist as children move to upper grades. McEwan and Shapiro (2008) in Chile and Crawford *et al*. (2007) in England find that maturity at enrolment gives older students a long-run advantage since relatively old children in each class still have higher scores in later grades than those born close to the cutoff date. Interestingly, a recent paper finds a fading differential as children move to higher grades [Elder and Lubotsky (2009)].

The age at first enrolment has also received independent attention. However, the majority of the existing studies, using changes in school policies or the legal entry age at school (cutoff date) as instruments to overcome endogeneity problems in their analyses, focus on entry into formal (compulsory) education [An-

grist and Krueger (1991), Elder and Lubotsky (2009)]. Research on differences in children's performance by age of entry into nursery school or less formal settings is more sparse. This is in part due to the difficulties of making causal inferences in these contexts without clean instruments for age of entry. We, nevertheless, consider that unveiling some basic associations between age at entry into any educational setting and later outcomes is an interesting question and one that we pursue here.

The two main findings of the paper are the following. First, quarter of birth is significantly related to cognitive development but not to social behavior and the maturity gap does not seem to dissipate as children advance into later grades. Even though our results cannot be interpreted in a causal manner, they indicate that cutoff rates deserve deeper study, as previous literature has already noted.

Second, the age of enrolment into any formal educational institution is related to children's development. Children who attended nursery school generally perform better than those who only start in the first year of preschool (*P3*) (three-year olds in Catalonia) or later. This relationship relaxes when origin and family education controls are included in the analysis. This suggests that those who start school late disproportionately belong to disadvantaged groups. Among those that attended nursery school, there are no large differences by age of entry.

Furthermore, we find a strong association between other demographic characteristics and educational outcomes. Among others, children raised in non-nuclear families, with low educated parents, with little exposure to after school activities or born outside Spain tend to underperform others at school. Finally, an other things being equal, children in public schools receive higher scores in cognitive knowledge than those in private schools with public funding (*concertada*). Given that school grades partly determine university access, the inability to determine whether this gap is driven by grade inflation in the public system confirms the need to use more standardised tests.

The paper is organised as follows. Section 1 summarises recent demographic trends in Catalonia and their relevance for education. Section 2 describes the data sources, the construction of the indicators of cognitive and non-cognitive knowledge, and the explanatory variables employed. Section 3 looks at the association between socioeconomic variables and development indicators using a multivariate analysis and discusses the results. Finally, the paper concludes with a summary of the findings in Section 4.

1. Recent demographic trends in Catalonia

In the last decade, Catalonia has undergone a large demographic transformation with more women entering the labour market, a massive immigration inflow and more children raised in adverse environments. These demographic changes have had a great impact on the structure of society, with more children attending nursery school earlier in life while mothers work and many immigrant children entering school with heterogeneous skills and educational backgrounds. In this context, it is crucial to identify the groups of children at risk of poor cognitive and non-cognitive development and to provide guidelines for enhancing their human capital.

Female participation rates (ages 16-64) rose from 55% in 1996 to 68% in 2007, a level above the average of the EU [Amarelo and Bové (2008)]. Higher maternal labour force participation is coupled with shifts to earlier ages of school enrolment and more prevalence of nursery school attendance. An extensive literature has focused on whether rising female employment rates may negatively impact the educational performance of children [Dronkers (1994)], and on the effects of age at school entrance on over different measures of school performance and, even, future labour market outcomes and health (see Stipek (2002) for a survey). This subject has made it into the mainstream press [see Weil (2007)]. Nevertheless, most of the emphasis has been on estimating the impact of entry age into formal education (i.e. kindergartern for the US) and not into any type of educational institution (i.e. nursery school).

We are interested in studying whether the age at which a child first enrolls into any form of school matters for his or her development in elementary school. Compulsory school in Catalonia starts in first grade, at age six. Primary (or elementary school) includes six grades and runs from age six to age twelve. However, pre-primary school, from age three to age six (*P3, P4, P5*), is recommended and provided at no cost by the public sector or at reduced fees by private schools that receive a public subsidy (*concertada*). As a result, the great majority of families choose to enroll their children at age three in *P3*. In addition, many families, particularly those with working mothers, send their children to nursery school before that age. The availability of public or subsidized nursery schools or childcare centres is somewhat limited. The fees for nursery schools, even public ones, are much higher than pre-primary and elementary schools. In Barcelona, for example, parents pay around 150 euros a month per child in a public nursery (0 to 3 years old) without including lunch while the cost is around 15 euros a month for 3 to 12 years old (just for supplies). In this paper, we control for the age of entrance into any form of school (nursery, pre-primary or elementary), not only compulsory education.

The number of foreigners registered in the census increased from 98,035 in 1996 to 972,507 in 2007. Migration contributed to 80% of the population growth in Catalonia during these years [Domingo and Bayona (2008)][1]. The inflow of foreigners was particularly large in the last five years, when more than half a million foreigners relocated to Catalonia. In fact, during this period, Spain received the largest net migration in absolute numbers among EU countries. The arrival of large flows of immigrants to Catalonia may have pulled the average age of entry in the opposite direction, that is, toward older ages. When they arrive in Catalonia, some immigrant children have either never before attended school or have a poor (or very heterogeneous) educational experience that requires remedial education in elementary and secondary school. Moreover, if a large number of students arrive once the academic year has started, they may disrupt the regular pace of the class and slow the progress of the rest of the class, leading to negative peer effects [Calero and Waisgrais (forthcoming) and Sánchez (2008)]. Studies that aim to estab-

(1) Population was 6,090,040 in 1996 and increased to 7,210,508 in 2007, according to the Census published in the Spanish Institute of Statistics.

lish best practices for assimilating these children into the school system [OECD (2003)] have proliferated with the new flows of migration.

Such competing demands pose problems for the policy maker as the allocation of resources across different stages of education is not trivial. Heckman and Masterov (2007) make a strong case for early intervention and propose that more resources be allocated to children of preschool age. They argue that investing in young children, especially those in disadvantaged environments, pays off. These individuals tend to become more productive and less dependent on future government benefits. However, at the same time, countries with large migrant populations need to devote resources to facilitate a smooth assimilation of the new population to overcome some initial handicaps.

Finally, data shows that the number of children raised in adverse environments has increased in modern economies. Heckman and Masterov (2007) refer to rising trends in different measures of adversity in the US (i.e. absence of a father, scarce financial resources and low parental education and ability). Catalonia and Spain follow similar trends. Flaquer (2008) notes, for example, that there has been a very sharp rise in divorce rates in Catalonia, which has led to a high number of monoparental families and left more children at risk of poverty. Nowadays, the divorce rate in Catalonia is 3.5 divorces per thousand inhabitants, well above the average in the EU, which stands at around 2 per thousand. Empirical evidence suggests that children born into disadvantaged families are likely to have worse future outcomes than others in areas such as education or employment [Jaffee *et al*. (2001)]. To address this, we control for the type of household in which children live and estimate its relationship to their educational performance and social behaviour.

2. Data

We use data from the project "Família i Educació a Catalunya" funded by the Fundació Jaume Bofill. The dataset contains socioeconomic information on 942 children aged 6 to 12. The random sample was constructed as follows. First, schools were randomly selected by stratum. Second, in each school, one of the three groups (2nd, 4th and 6th of Primary) was randomly selected. Finally, all children from that grade were selected. More details about sample construction can be found in Bonillo *et al*. (2007).

The information for each child was collected in 2005 through four questionnaires: two completed by the parents (one with general household information and another about the children's social attitudes); a third completed by the teacher about the children's cognitive and non-cognitive abilities at school as well as other social behaviour characteristics; and a fourth (not used in this paper) filled in by children aged 12 years old about their social attitudes.

In the household information survey, parents were asked about the structure of the family (origin, education, labour force status, parental age), language used at home, household income, school factors (age of enrolment, reasons for the selection of a particular school and degree of participation of the parents in meetings), out-of-school activities, resources and expenditure in education, norms, values and religious and political attitudes.

The second questionnaire completed by parents included three main sections. The first one dealt with the social behaviour of the child and follows the Strengths and Difficulties Questionnaire (SDQ) in Goodman (1997) (http://www.sdqinfo.com). These questions were also answered by the teachers in their questionnaire. A second set of questions discussed the relationship between the child and his/her family (i.e. interaction and conversation) and followed the structure of the Alabama Parenting Questionnaire (APQ) in Frick (1991) (http://fs.uno.edu/pfrick/APQ.html). The final section revolved around aspects of family life such as whether there are lively discussions and the members of the household have a good time together. This part followed the Self-report Family Inventory (SFI) in Beavers *et al*. (1990).

Teachers completed a survey with three sections. First they assessed the cognitive knowledge of the student in seven subjects and in the Catalan language. Second, they evaluate the qualitative school abilities of the child such as whether the child undertakes school tasks in an adequate manner (i.e. delivered on time, asking for help when needed, etc.), whether he/she follows the general rules and pays attention to the teacher, and whether he/she is well integrated into the group and collaborates with other pupils. This section followed the *School Social Behaviour Scales* (SSBS-2) in Merrell (2002). The third section covered the same questions about social behaviour as the parents were assed and was based on the SDQ in Goodman (1997).

2.1. Cognitive and non-cognitive indicators

A complete picture of a child's educational performance requires a multifacet view of the child's development. Both academic performance and emotional and social maturity are important to evaluate a child's achievement. Substantial research shows that, for example, employment success clearly depends on both cognitive and non-cognitive abilities [Heckman *et al*. (2006)]. This paper uses four indicators to measure children's development: two for cognitive learning (*Global knowledge* and *Catalan Knowledge*) and two for non-cognitive behaviour (*School Abilities* and *Social Behaviour*).

The two variables for cognitive learning are constructed as follows. *Global knowledge* assesses the general academic knowledge of the child and is calculated by taking an arithmetic average of the teacher evaluations (on a scale of 1 to 3: low, average or high) on seven subjects (Science, Catalan, Spanish, Foreign language, Mathematics, Art and Physical education). *Catalan Knowledge* assesses the child's knowledge of the language and is an arithmetic average of writing skills, reading, oral expression and comprehension, each evaluated on a scale of none (0) to very high (10) (see Appendix B for details).

The first indicator for non-cognitive knowledge, *School Abilities,* is the average of twenty qualitative responses (on a scale from 1 to 5) given by the teachers on school competence of the children. They include information on the academic performance of the pupil, his/her self-control and ability and willingness to follow directions in the classroom as well as his/her personal relationship with other students at school.

The second indicator, *Social Behaviour*, is the average of the parents' and teacher's answers on a scale from 1 to 3. Taking the mean of two different sources should provide a more objective (or balanced) measure of a child's social behaviour. Sample descriptives show that parents' responses tend to be, in general, more

positive than those of teachers. To see whether this biases our inferences in any way, we calculate an alternative indicator that only uses the least favourable of the two values, either the parents' or the teachers' scores, for each student as their evaluation. This is a strategy often recommended when indicators are constructed from responses given by respondents who may overstate their answers [see, for example, Piancentini *et al*. (1992)]. Our results, however, are robust to the use of either indicator and in the text we only present estimates with the indicator calculated as an average of the teacher's and the parent's evaluations.

The parents' and teachers' indices, in turn, have been calculated from the responses to several questions in SDQ on social behaviour such as the kindness or temperament of the child. Some of the questions in SDQ are positively phrased while others are negative. Moreover, responses to some questions are interrelated and need to be combined to make sense. Given these characteristics of the SDQ survey a simple average of the responses (as is done for the *School Abilities* based on SSBS-2) does not provide the appropriate final indicator. Instead, it is necessary to use the algorithm to extract the scores from the responses of the survey and classify them into three levels "normal", "borderline", and "abnormal" (see Goodman (1997) for guidelines).

Table 1 presents the means and standard deviations of these indicators. Average scores in the sample are the following: 2.3 for *Global Knowledge* (scores range from 1 to 3), 7.1 for *Catalan Knowledge* (scores range from 0 to 10), 3.9 for *School Abilities* and 2.8 for *Social Behaviour* (scores range from 1 to 3).

Table 1: COGNITIVE AND NON-COGNITIVE INDICATORS

Variables	Observations	Mean	Std. Dev.
Global Knowledge	928	2.326	0.519
Catalan Knowledge	937	7.137	2.018
School Abilities	919	3.903	0.721
Social Behaviour	927	2.819	0.359

Source: Own elaboration.

2.2. Independent variables

Quarter of birth

The analysis includes information on the quarter of birth: *Birth quarter 1* from January to March; *Birth quarter 2* from April to June; *Birth quarter 3* from July to September; *Birth quarter 4* from October to December. December 31st is the school cutoff date in Catalonia (and Spain), and children born later in the year will be the youngest in the class.

Age of first attending any form of school

We control for the age when a child first attended any form of school by including the following dummies: <=1, =1.5-2, =2.5-3.5, =4-5.5, >=6-7. The refer-

ence group is before age 1.5. Even though compulsory education does not start until age six, school is provided publicly (or heavily subsidized in the private sector) from age three (*P3*). As a result, a great majority of children enter in *P3*. Given that the cutoff date for children in Catalonia is December 31st, children who are entering school for the first time before turning three but who were born between September and December are attending formal preschool (*P3*) and not nursery school before turning three. For this reason, we use an additional specification that includes three dummies for the first type of school attended: *Nursery* if the child went to nursery school before starting free preschool (*P3*) at age 3; *Preschool (P3)* if the child started at *P3* ; *After P3* if the child started after *P3*. This variable has been constructed by combining the age of first school attendance and the month of birth, and by taking into account that the academic year starts in September.

Personal and family characteristics

The models include a rich set of personal information: Age when the interview takes place (and its square); gender (*Girl* 1 for a girl and 0 for a boy); three dummies for the grade at which the child is currently enrolled in primary school (2^{nd} *Grade* for second, 4^{th} *Grade* for fourth and 6^{th} *Grade* for sixth). Since we control for quarter of birth and grade, the interpretation of the variable *Age* is unclear. It combines both the effect of being held (most likely because of poor performance since parents in Catalonia do not have a choice to enroll their children later than the legal age) and of being the eldest within the same quarter of birth.

The analysis also contains information on the composition of the family: the number of siblings (*N.Siblings*); four dummies for the ranking of birth (*Single child* if the child does not have any brothers or sisters, *Eldest child* if the child is the eldest sibling, *Middle child* if the child is between siblings, *Youngest child* if the child is the youngest sibling); four dummies on the structure of the family including *Nuclear* (both parents and children live together in the same household), *Monoparental* (only one of the parents lives with the child), *Extensive* (children cohabit with other members of the family such as the grandparents in addition to their parents) and *Rebuilt* (children also cohabit with individuals that are not members of their biological or adoptive family such as a step-parent).

Regarding the socioeconomic characteristics of the family, we include six dummies for origin of the child: *Catalonia*, *Rest of Spain*, *Eu-USA-Aus-NZ* for Europe, North America, Australia and New Zealand, *Asia*, *Latin America* and *Africa*. In parallel regressions, we add four dummies for the origin of the parents to look at the effect of second generation immigration: both are Catalan, one of them is Catalan, one is from the rest of Spain (but none is Catalan); both are born out of Spain.

We also include information about the language spoken at home: *Catalan* if only Catalan, *Spanish* if only Spanish, *Cat. or Span.* if some Catalan or Spanish and *Other* if neither Catalan nor Spanish.

Three dummies summarise overall family education: *Lower than Secondary High* which encompasses low secondary or less, *Secondary High* for upper secondary and *University*.

The labour force status of the parents is controlled: None employed if neither the father nor the mother is employed, *One employed* if either the father or the mother is employed and *Both employed* if both father and mother are employed.

Three dummies account for the net monthly income of the household: *<= 1,800, 1,800-3,000 and > 3,000.* We imputed the average income to the 45 observations with missing income information and added a dummy variable in the regression to control for those cases [Cohen *et al*. (2003)].

Finally, we control for the religious involvement of the child: *Relig./Pract* if the child is religious and practising, *Relig./not Pract.* if the child is religious but does not practise and *Not Religious* otherwise.

School characteristics

In the regressions, we add school fixed effects and other school characteristics: four dummies for the size of their municipality (*>500,000*, *50,000-500,000*, *5,000-50,000*, *<5,000*); a dummy variable (*Public School*) for public or private with some public funding (*concertada*); three dummies for the total size of the school, each of them containing roughly one third of the sample (<= 300, *300-600* and > 600). In parallel regressions, we also included the number of students per class, the proportion of students in a class who either enroll once the academic year has started or with special needs. However, we excluded these variables from the final analysis since around 65 individuals reported that their class size was larger than 30, which appears inconsistent with the legislation in place. Estimates for these regressors were not significant.

Time use by the student

Estimates include information on the total number of hours per week that a each child devotes to activities that are likely to affect his/her school and social performance: more academic after-school activities such as language, computer science and music (*Intellectual*); sport related after-school activities such as dance, sports and psychomotor activity (*Sports*); *Reading*; doing *Homework*; and watching *TV*.

2.3. Description of the sample

Table 2 displays the means and standard deviations of the independent variables. The students in our sample are evenly divided across gender, grades (second, fourth and sixth) and quarter of birth. Regarding when children first entered some type of school, roughly a third of them did so before their first birthday. Almost half of them had enrolled by eighteen months and three quarters by the time they were two and a half years old. Only 5% of the sample enrolled in a school for the first time after age three. Distinguishing between nursery and preschool, we observe than almost three quarters of the children went to nursery school, around 17% started around age 3 at P*3* and the remaining 10% enrolled later. Nursery attendance rates do not match those from the Ministery of Education for the academic year 2006-2007 ("Las cifras de la educación en España"). While the Ministery reports that the net schooling rate in authorised centres for 2 year olds was 33%, our sample displays a substantially higher rate (based on the second half of the 90s). Data on preschool rates are more in line with our sample (according to the Ministery, the schooling rate for 3 year olds was 96.7%). Therefore, overall, our sample seems to be overestimating the proportion of children who attend nursery before the age of 3. This disparity in schooling rates may partly reflect differences in sampling and classifcations such as age groups and type of schools.

Table 2: DESCRIPTIVE STATISTICS

Variables	Observations	Mean	Std. Dev.
Girl	942	0.491	0.500
Age	940	9.652	1.660
Quarter of Birth			
Birth q. 1	942	0.262	0.440
Birth q. 2	942	0.251	0.434
Birth q. 3	942	0.220	0.415
Birth q. 4	942	0.263	0.440
Grade			
2nd Grade	942	0.336	0.472
4th Grade	942	0.371	0.483
6th Grade	942	0.291	0.454
Age Start School			
<=1	942	0.342	0.474
=1.5-2	942	0.374	0.48
=2.5-3.5	942	0.238	0.426
=4-5.5	942	0.031	0.175
>=6-7	942	0.010	0.102
Nursery	942	0.736	0.440
P3	942	0.165	0.371
Start after P3	942	0.097	0.297
Family Type			
Nuclear	941	0.790	0.407
Monoparental	941	0.106	0.308
Extensive	941	0.057	0.232
Rebuilt	941	0.045	0.208
N. Siblings	941	1.031	0.722
Single Child	940	0.200	0.400
Eldest Child	940	0.385	0.486
Middle Child	940	0.074	0.262
Youngest Child	940	0.340	0.474
Birthplace			
Catalonia	940	0.892	0.309
Rest of Spain	940	0.019	0.137
Eu-USA-Aus-NZ	940	0.008	0.091

Table 2: DESCRIPTIVE STATISTICS (continuation)

Variables	Observations	Mean	Std. Dev.
Asia	940	0.003	0.056
Latin America	940	0.067	0.250
Africa	940	0.009	0.097
Parents' Birthplace			
Both Catalan	941	0.596	0.490
One Catalan	941	0.217	0.413
At least one Spanish (none Catalan)	941	0.086	0.280
Out of Spain	941	0.099	0.300
Home Language			
Catalan	941	0.523	0.499
Spanish	941	0.361	0.480
Catalan and Spanish	941	0.102	0.302
Family Education			
None/Primary	937	0.001	0.032
Secondary Low	937	0.149	0.356
Secondary High	937	0.406	0.491
University	937	0.439	0.496
Parental Employment			
None employed	940	0.021	0.144
One employed	940	0.299	0.458
Both employed	940	0.679	0.467
Monthly Net Family Income			
< 600	942	0.007	0.085
600-1,800	942	0.312	0.463
1,800-3,000	942	0.423	0.494
> 3,000	942	0.256	0.437
Missing	942	0.047	0.213
Child Religion			
Religious/Practice	942	0.215	0.411
Religious/not practice	942	0.402	0.490
Not Religious	942	0.382	0.486
Municipality Size			
> 500,000	942	0.212	0.409
50,000–500,000	942	0.321	0.467
5,000–50,000	942	0.357	0.479
< 5,000	942	0.108	0.310

Table 2: Descriptive statistics (continuation)

Variables	Observations	Mean	Std. Dev.
Type of School			
Public School	942	0.636	0.481
Number of students in the school			
< 300 Students	935	0.294	0.455
300-600 Students	935	0.372	0.483
> 600 Students	935	0.333	0.471
Hours a week in after-school activities			
Intellectual	942	1.203	1.775
Sports	942	2.871	2.728
Reading	942	2.634	2.157
Homework	942	5.138	3.550
TV	942	9.360	6.490

Source: Own elaboration.

With regard to the structure of the family, 20% of the children in the sample are single children, 60% have one sibling, 17% have two and only 3% have three or more siblings. Around 80% of the families are nuclear and 10% monoparental. The remaining 10% is roughly equally divided between extensive and rebuilt families.

In our sample, close to 10% of the students were born outside Catalonia: 2% in the rest of Spain, 7% in Latin America, 1% in Maghreb and 1% in other countries. Among students, 10% were second generation immigrants with both parents born outside Spain. Regarding the language regularly spoken at home, 52% speak only Catalan at home, 36% speak only Spanish, 10% mix Spanish and Catalan at home and the remaining 2% does not speak either Catalan or Spanish.

3. Socioeconomic factors and education attainment

This section analyses the association between socioeconomic characteristics and the level of cognitive and non-cognitive knowledge, which is measured with the four indicators described in Section 3. For this purpose, we estimate a multivariate model with ordinary least squares (OLS) including a large set of explanatory variables.

All four dependent variables (*Global knowledge, Catalan Knowledge, School Abilities* and *Social Behaviour*) are derived from information given by the teacher and not from an external objective examination as is the case, for example, in the Programme for International Students Assessment (PISA) [OECD (2006)]. Therefore, it is possible that teachers from certain schools tend to give higher marks (grade inflation) than those from other centres. For this reason, it is important to take into account that the educational data on students have been sampled from

many schools and each school is a cluster. Outcomes within a cluster are likely to be correlated and for this reason we include school fixed effects in the regression. Results of the multivariate analysis are shown in Table 3.

Table 3: PARAMETER ESTIMATES OF COGNITIVE AND NON-COGNITIVE KNOWLEDGE

Variables	Global Knowledge	Catalan Knowledge	School Abilities	Social Behaviour
Girl	0.083**	0.380***	0.273***	0.155***
	(0.039)	(0.142)	(0.054)	(0.026)
Age	0.198	0.748	-0.069	0.123
	(0.252)	(0.923)	(0.339)	(0.221)
Age square	-0.012	-0.045	-0.001	-0.004
	(0.013)	(0.046)	(0.017)	(0.012)
Quarter of Birth (Birth quarter 1, omitted)				
Birth q. 2	-0.124**	-0.224	-0.124*	0.001
	(0.052)	(0.192)	(0.074)	(0.037)
Birth q. 3	-0.193***	-0.617***	-0.232***	0.008
	(0.058)	(0.211)	(0.079)	(0.041)
Birth q. 4	-0.275***	-0.778***	-0.360***	0.002
	(0.065)	(0.247)	(0.090)	(0.044)
Grade (2nd Grade, omitted)				
4th Grade	0.507*	6.716***	-0.341	1.071***
	(0.280)	(1.048)	(0.354)	(1.164)
6th Grade	0.413	7.267***	-0.916**	0.936***
	(0.294)	(1.031)	(0.368)	(0.188)
Age Start School (<=1, omitted)				
=1.5-2	0.073*	0.339**	0.088	0.005
	(0.043)	(0.159)	(0.061)	(0.032)
=2.5-3.5	-0.047	0.124	0.101	0.002
	(0.055)	(0.198)	(0.075)	(0.038)
=4-5.5	-0.095	-0.299	0.048	0.164
	(0.127)	(0.601)	(0.209)	(0.119)
>=6-7	-0.411**	-1.445**	-0.430*	0.067
	(0.145)	(0.675)	(0.245)	(0.158)

Table 3: Parameter estimates of cognitive and non-cognitive knowledge (continuation)

Variables	Global Knowledge	Catalan Knowledge	School Abilities	Social Behaviour
Family Type (Nuclear, omitted)				
Monoparental	-0.012	-0.169	-0.084	-0.041
	(0.077)	(0.262)	(0.101)	(0.056)
Extensive	-0.074	-0.500*	0.033	-0.026
	(0.072)	(0.302)	(0.103)	(0.063)
Rebuilt	-0.268**	-0.439	-0.237*	0.029
	(0.107)	(0.340)	(0.129)	(0.075)
N. Siblings	-0.038	-0.099	-0.031	-0.024
	(0.034)	(0.141)	(0.048)	(0.026)
Birth Order in the Family (Middle/Youngest, omitted)				
Single Child	0.039	0.237	-0.034	-0.060
	(0.070)	(0.266)	(0.105)	(0.052)
Eldest Child	0.061	0.357**	0.011)	-0.041
	(0.043)	(0.160)	(0.060	(0.030)
Birthplace (Africa, omitted)				
Catalonia	0.454**	1.958*	1.032***	0.462*
	(0.216)	(1.070)	(0.301)	(0.246)
Rest of Spain	0.427*	1.871	0.905**	0.418
	(0.251)	(1.154)	(0.372)	(0.273)
Eu-USA-Aus-NZ	0.027	0.229	0.744*	0.592**
	(0.291)	(1.216)	(0.410)	(0.264)
Asia	0.598***	1.501	1.355***	0.424
	(0.227)	(1.124)	(0.466)	(0.314)
Latin America	0.274	0.888	1.018***	0.480*
	(0.214)	(1.066)	(0.305)	(0.248)
Home Language (Other, omitted)				
Catalan	0.018	0.124	0.066	0.103
	(0.109)	(0.454)	(0.176)	(0.089)
Spanish	0.080)	-0.020	-0.014	0.016
	(0.109	(0.462)	(0.173)	(0.088)

Table 3: Parameter estimates of cognitive and non-cognitive knowledge (continuation)

Variables	Global Knowledge	Catalan Knowledge	School Abilities	Social Behaviour
Cat. and Span.	0.089	0.224	0.105	0.095
	(0.114)	(0.483)	(0.184)	(0.097)
Monthly Net Family Income (< 1,800, omitted)				
1,800-3,000	0.143***	0.454**	0.104	0.035
	(0.051)	(0.190)	(0.071)	(0.038)
> 3,000	0.103*	0.363	0.039	0.079*
	(0.063)	(0.228)	(0.089)	(0.042)
Missing	-0.194*	-0.555*	-0.306***	-0.104*
	(0.099)	(0.314)	(0.115)	(0.060)
Family Education (Lower than Secondary High, omitted)				
Secondary High	0.044	0.103	-0.094	-0.066
	(0.058)	(0.226)	(0.083)	(0.045)
University	0.198***	0.668**	0.035	-0.080
	(0.068)	(0.260)	(0.092)	(0.053)
Parental Employment (None Employed, omitted)				
Both Employed	-0.260	-0.914*	-0.227	0.097
	(0.168)	(0.497)	(0.206)	(0.112)
One Employed	-0.223	-0.618	-0.272	0.063
	(0.167)	(0.490)	(0.202)	(0.108)
Child Religion (Not religious, omitted)				
Relig./Pract.	0.063	0.249	0.001	-0.009
	(0.052)	(0.198)	(0.069)	(0.035)
Relig./not Pract.	0.107**	0.398**	-0.041	-0.091***
	(0.045)	(0.165)	(0.062)	(0.032)
Municipality Size (> 500,000, omitted)				
50,000-500,000	0.721***	1.690*	1.610***	0.264*
	(0.249)	(0.937)	(0.298)	(0.136)
5,000-50,000	0.364	0.151	1.058***	0.318***
	(0.278)	(0.843)	(0.317)	(0.111)
< 5,000	0.740**	3.395***	-1.671***	0.554***
	(0.308)	(0.925)	(0.298)	(0.154)

Table 3: Parameter estimates of cognitive and non-cognitive knowledge (continuation)

Variables	Global Knowledge	Catalan Knowledge	School Abilities	Social Behaviour
School Characteristics (Semi-private, omitted)				
Public School	0.331*	0.919	-0.097	-0.211**
	(0.199)	(0.885)	(0.277)	(0.104)
Number of students in the school (< 300, omitted)				
300-600 students	-0.746***	-3.543***	-1.622***	-0.240**
	(0.174)	(0.666)	(0.226)	(0.115)
> 600	-0.131	1.227	-0.516**	-0.406***
	(0.237)	(1.053)	(0.260)	(0.150)
Hours a week in after-school activities				
Intellectual	0.040***	0.172***	0.022	-0.006
	(0.012)	(0.041)	(0.015)	(0.008)
Sports	0.020***	0.075**	0.011	0.006
	(0.007)	(0.029)	(0.010)	(0.006)
Reading	0.036***	0.137***	0.034**	0.004
	(0.009)	(0.031)	(0.013)	(0.006)
Homework	-0.008	-0.050**	-0.010	0.005
	(0.007)	(0.024)	(0.009)	(0.005)
TV	0.000	0.005	-0.000	-0.002
	(0.003)	(0.012)	(0.004)	(0.002)
Constant	0.458	-4.742	3.505**	2.906***
	(1.365)	(4.661)	(1.783)	(1.058)
Observations	913	922	904	912
R^2	0.436	0.474	0.440	0.343

*** p < 0.01, ** p < 0.05, * p < 0.1. School-fixed effects. Robust standard errors in parentheses.

Source: Own elaboration.

The variable *Global knowledge* is a mix of seven subjects, including physical education, which probably does not really qualify as a cognitive skill. As a robustness check, we have also estimated models with alternative definitions of cognitive knowledge focused on particular skills such as mathematics, science and languages, and find similar results.

Another characteristics of our dependent variables is that their range of values is limited, as shown in Table 1. The fact that the average of *Global Knowledge* is high (2.3) and has a standard deviation of 0.5 suggests that a non-trivial number of observations are located at the maximum value (3). In order to observe how this may affect the outcome of our regressions, we have also estimated a Tobit model to account for the upper and lower bounds of the dependent variables, and find comparable results.

3.1. Quarter of birth

Results in Table 3 suggest that there are substantial differences in performance according to when a child is born. Ceteris paribus, the scores of children born at the beginning of the academic year are in general higher in *Global Knowledge*, *Catalan Knowledge* and in *School Abilities*. Children born at the end of the year may have an initial disadvantage compared to their classmates. In separate estimates that include interactions of quarter of birth with each grade level, we find some evidence that the maturity gap does not diminish with age. Overall, if anything, it seems that the negative gap increases for those born in the last quarter of the year though none of the coefficients is significant. Interestingly, a recent paper finds a fading differential as children move to higher grades [Elder and Lubotsky (2009)]. Conversely, McEwan and Shapiro (2008) find evidence in Chile that maturity at enrolment gives older students a long-run advantage since relatively old children in each class still have higher scores in fourth and eighth grade than those born close to the cutoff date. The authors claim that the persistent effects suggest that older enrolment age rather than age-at-test explain the better performance of these students. Along these lines, Crawford *et al*. (2007) find that the month in which you are born matters for test scores at ages 7, 11, 14 and 16 in England, with younger children performing significantly worse, on average, than their older peers. The authors are able to exploit the geographic differences within England of the length of schooling and the age at which children start school to identify these effects. Unfortunately, in our case, the school-entry policies are homogenous across all schools and we cannot construct any type of instrumental variable as other authors have done. Our results need to be read within that framework.

With regard to particular subjects, Table 4 shows that children born late in the year tend to underperform others in all areas except for physical education. The negative estimate is larger in Spanish, Foreign Language and Mathematics, which suggests that compensatory programs, if implemented, should probably focus more carefully on these subjects.

One source of concern for interpreting the findings on quarter of birth is that parents with different socioeconomic characteristics and ability to access family planning will choose to have children at different seasons of the year (i.e. either earlier in the year when there are better prospects of succeeding at school, or at the end of the spring to enjoy longer holidays or to have infants during the months of better weather). If there was indeed differential parental planning, our results could be biased. However, there are good reasons to think that this is unlikely since there are many obstacles to perfect planning and, in most cases, conception does not occur necessarily in the first cycle when it is attempted. Our data shows that the frequency of births during the year is pretty much constant.

Table 4: Parameter estimates by subjects

	Science	Catalan	Spanish	Foreign L.	Maths	Arts	Gym
Birth q. 2	-0.119	-0.103	-0.108	-0.167*	-0.110	-0.129	-0.088
	(1.80)	(1.39)	(1.61)	(2.58)	(1.55)	(1.89)	(1.43)
Birth q. 3	-0.195*	-0.203*	-0.246**	-0.240**	-0.244**	-0.19*	-0.104
	(2.53)	(2.47)	(3.25)	(3.16)	(2.81)	(2.53)	(1.45)
Birth q. 4	-0.256**	-0.282**	-0.382**	-0.370**	-0.323**	-0.275**	-0.139
	(2.97)	(2.96)	(4.60)	(4.53)	(3.25)	(3.30)	(1.73)
Observations	921	921	919	900	917	908	892

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. School fixed effects. Robust standard errors in parentheses.

Same regressors as in Table 3.

Source: Own elaboration.

One way to check whether parental planning plays a significant role in our study would be to compare children born during the last and first days of the year to see whether being the youngest or oldest in a school grade matters for performance. The rationale behind this is that planning for an exact day or week of birth may be more difficult than generally planning for a "Spring" versus a "Fall" baby. We would like to use the discontinuity provided by the cutoff date of January 1st to see whether there is any difference in performance. For example, we would like to observe children born at the end of December currently in 4th grade versus those born in early January currently in 3rd grade. Unfortunately, our dataset does not provide continuous information of children from 2nd grade to 6th grade, but jumps two grades each time, which means that we do not have the natural comparison group to undertake such an analysis.

Because of this data limitation, we undertake a cruder check that consists of exploring the differences in the various measures of development between children born in January and December in the same year. We expect the parents of these children to be relatively similar in terms of planning their births in a particular season though, of course, making the decision a year later and that, if quarter of birth matters, the difference in outcomes should be greatest when restricting attention to these two groups. The estimated coefficients of this exercise point to large differences in children's cognitive development between those born in the first and last months of the year. However, none of the coefficients attains significance at reasonable levels if school fixed effects are included. The coefficient for the month of January (compared to the benchmark of a December birth) on the indicator on social attitudes is very small and insignificant. This is in line with our a priori expectation that the differences should not be too large, if any, in that sphere. The fact that the coefficients are not statistically significant can be probably explained by the small number of observations left when restricting the analysis to children born in January and December. Overall, our results are in line with

those of the quarter of birth on children's cognitive development though with the current sample size we cannot make a strong case[2].

A second problem with interpreting the coefficients of the quarter of birth may arise if the quarter of birth is endogenous to some parental characteristics (i.e. age, marital status and education of the mother) that may account for part of the differential performance [Bound and Jaeger (2000), Buckles and Hungerman (2010)]. This may bias the results. There are indeed recent studies for the US that find that children born in winter months are disproportionately born to women who are teenagers, have less than high school education and who are unmarried [Buckles and Hungerman (2010)]. In order to check this is our case too, we undertake a multinomial logit analysis where the dependent variable is the quarter of birth and the independent variables are the socioeconomic characteristics of the parents (e.g. age, education, birthplace, labour market status, city of residence). Results, available upon request, show that there is no correlation between the parental socioeconomic characteristics and the quarter of birth in our data.

Finally, there is a third source of concern related to the month of birth of the children: the type of daycare centre attended (public and private). Registration to attend public daycare centres for the academic year starting in the September of a given year requires that the child be born by the end of April of that year. Otherwise, the child has to wait until the next academic year. As a result, it is possible that more children born at the beginning of the year will end up in public daycare centres while their younger counterparts will enroll in the private nursery schools and will probably stay there until starting preschool in *P3*. In this case, if there are differences in quality between public and private nurseries, and the allocation between the two types of school depends on the month of birth, then estimates for the quarter of birth may be also picking up the type of nursery attended and not only the effect of the month of birth itself. Unfortunately, this information is not available in the sample and, for the moment, the only thing we can do is to control for the season of birth and whether the child was or not enrolled in preschool.

On these grounds, we carry out a parallel exercise where we limit our sample to children born in April (eligible for public daycare) and May (not eligible). The rationale behind this is that, if we find large differences in development between children born in these two consecutive months, then the type of school attended may be an important determinant for children's future development. Results (available upon request) show that the differences between children born in April and May are very small. If anything, they seem to be in favour of those born in May. However, since we do not have any means to see what type of childcare the kids born in April attended, we cannot infer from the results that children who attend public childcare tend to do worse later on. Furthermore, the alternatives to attending public daycare chosen by those born in May are di-

(2) Results from these parallel regressions, not shown here, are available upon request. When excluding school fixed effects, the coefficient for a January birth is positive and significant for the first three indicators.

verse, either attending private daycare or staying longer at home with a parent, grandparent or babysitter[3].

It is important to note that using season of birth on its own to try to capture the quality of daycare centres may be confounding with "length of time in daycare". For example: a child born in early January and not able to be enrolled in *P3* until 3 years and 9 months old, would have stayed longer in preschool than a child born in December and enrolled in *P3* at 2 years and 9 months old.

The fact that the quarter of birth matters for educational attainment raises questions about whether the allocation of children across academic years should be different and whether some policies could be implemented to help children who are relatively young at school enrolment. For instance, Bedard and Dhuey (2006) discuss the possibility of grouping students by ability. Some researchers have carried out randomized experiments to address this question. Cascio and Schanzenbach (2007) use data from one of the largest educational experiments ever carried out in the US (STAR project) to study whether relative age in a classroom matters. For this project, children of the same biological age were randomly assigned to different classrooms at school entry. They find no evidence that relative age matters for the average student, but disadvantaged students, when placed among older children in the same classroom, are less likely to take a college entrance exam than other kids of their exact same age.

An alternative policy would involve compensatory programs like remedial tutoring for lower-achieving, relatively young students [Chay *et al*. (2005)]. Another option would be to increase the number of grades so that the age spread between the children would be smaller. Interestingly, the gap in children's performance between birth quarters only appears in academic performance measures, but not for *Social Behaviour*, a non-cognitive indicator.

3.2. Age when child first attended any form of school

Regarding the age when a child first attended any form of school, there is a strong negative correlation between starting school after turning two and half and cognitive knowledge. This association persists once other personal variables such as age, quarter of birth, grade level and the composition of the family are added to the regression (see Table 5). However, when information about origin is included, some coefficients of age at the time of first enrolment are only marginally significant and their significance disappears except for those who start at age six or later, once variables such as income and educational background of the parents are added to the multivariate analysis. This suggests that children starting late generally do worse partly because they share background characteristics that are strongly negatively related to general academic performance (i.e. born out of Catalonia, belong to a socioeconomic disadvantaged group).

This interesting finding matches some previous results in the literature. Cascio and Schanzenbach (2007) find that disadvantaged children who are older at

(3) We also estimated models with interactive variables of the different ages at first school attendance and the month of April and found no advantage among those born in April and attending day care very early on.

Table 5: Parameter estimates of global knowledge

	All				Boys	Girls
Variables	(1)	(2)	(3)	(4)	(5)	(6)
Age Start School (<=1, omitted)						
=1.5-2	0.043	0.031	0.043	0.073*	0.107	-0.065
	(0.044)	(0.044)	(0.045)	(0.043)	(0.068)	(0.082)
=2.5-3.5	-0.116**	-0.119**	-0.096*	-0.047	-0.073	-0.013
	(0.054)	(0.055)	(0.056)	(0.055)	(0.082)	(0.114)
=4-5.5	-0.410***	-0.418***	-0.221*	-0.095	-0.390*	0.047
	(0.113)	(0.114)	(0.131)	(0.127)	(0.0.211)	(0.230)
>=6-7	-0.761***	-0.714***	-0.463***	-0.411***	-0.484**	-0.312
	(0.144)	(0.143)	(0.156)	(0.145)	(0.186)	(0.282)
Observations	925	925	924	913	451	462

*** p < 0.01, ** p < 0.05, * p < 0.1. School fixed effects. Robust standard errors in parentheses. The number of explanatory variables increases from model (1) to (4).

(1) Controls for gender, age, quarter of birth and grade.

(2) Adds family type, number of siblings and birth order in the family.

(3) Adds birth place and home language. (4), (5) and (6) All variables as in Table 3.

Source: Own elaboration.

the start of kindergarten are less likely to take college entrance exams, but the opposite is true for children from high socioeconomic backgrounds. Similarly, Elder and Lubotsky (2009) note that the differences researchers find by birth quarter in elementary school may be related to the previous experience of the children in nursery school (prior to kindergarten in the US; or prior to P*3* in this context). Children who are born earlier in the year may have had a much longer exposure to preschool and nursery environment before going into first grade than those who enter formal education half a year younger. Furthermore, even though Elder and Lubotsky (2009) find that the differences in test scores according to entrance age decline sharply past kindergarten, they are especially large for children with a more advantageous background. That is, children who are older at school entry and who, given their family economic background, may have been exposed to better or longer quality nursery/preschool activities than those of a lower socioecomomic background do much better (particularly during the first years). Increasing access to preschool and nursery schools for more disadvantaged families is a potential policy implication of these findings. Table 5 also shows that enrolling late in school tends to reduce academic performance, especially for boys.

Moreover, we do not observe a significant difference in general performance and social attitudes between children who went to nursery school before turning one and a half and those who started slightly later between one and a half and two years

of age. Coefficient estimates indicate that 1 to 2 is probably the best age of entry. Not surprisingly, children who begin school at age six or later score much lower in *Catalan Knowledge*. This suggests that bringing children to school at an early age improves their lifetime language knowledge. The result, though, is partly driven by the fact that most of the children starting school that late are newcomers into the country.

Alternatively, we have divided the children into three broad groups of 'academic' initiation: nursery, first year of preschool (*P3*) and after. Results in Table 6 show that, ceteris paribus, children who went to nursery school have higher cognitive knowledge than those who started later, and the estimates are statistically significant for our *General Knowledge* indicator. Hence, overall, there seem to be some academic gains from attending nursery school. Currently, resources devoted to preschool in Spain are much lower than those for primary, secondary and university. Public expenditure in nurseries (0 to 3 years old) is 0.1% of the GDP. The rates are 0.5%, 1.1%, 1.7% and 1% for pre-primary (3 to 6), primary, secondary and tertiary school, respectively. Our findings suggest that it may be more efficient to distribute investment in education differently.

Table 6: Parameter estimates of cognitive and non-cognitive knowledge

Variables	Global Knowledge	Catalan Knowledge	School Abilities	Social Behaviour
Age Start School (*Nursery*, omitted)				
Preschool (P3)	-0.107**	-0.260	0.002	-0.014
	(0.054)	(0.198)	(0.073)	(0.038)
After P3	-0.130*	-0.128	0.049	0.084
	(0.077)	(0.299)	(0.119)	(0.054)

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. School fixed effects. Robust standard errors in parentheses. Same regressors as in Table 3.

Source: Own elaboration.

At any rate, these results should be interpreted with caution. Age of entrance to formal schooling is directly affected by parental decisions and if available family controls do not cover all the relevant information we may run into an omitted variable problem. Note, however, that this finding persists even after we control for family income among a battery of controls to proxy for some unobservable characteristics that could affect performance positively. To address endogeneity problems of entry age and be able to make causal inferences many studies have employed the quarter of birth or the legal entry age at school (cutoff date) as an instrument for entry age [Angrist and Krueger (1991), Elder and Lubotsky (2009)]. Results have been mixed though, in general, they provide some evidence of better performance among the older members of a class, even though some of these works find that these differences fade over time.

One alternative solution would be to instrument age of entry with a measure of "predicted" age of entry, which would be a student's entry if they had entered strictly according to the elementary school admission laws. However, this method would not work in the Catalan context because very few students actually start compulsory school at age six. Instead, parents exercise a lot of discretion and most students start during nursery or preschool years well before the legal age of six. Hence, a measure for predicted age of entry would lack the strong relationship needed between predicted and actual age of entry for the instrument to work. It is very difficult to find an appropriate instrument for age of entry. One possibility may be to use the degree of involvement of grandparents (frequency of visits) in their grandchildren's care to proxy for alternatives to sending children to nursery school early in life[4]. We use current grandparental care as a lower bound proxy since we do not observe this information when the child was an infant or at the precise time when a child first started school. Results, however, reject the validity of this instrument and this variable is insignificant when added to the main regression.

3.3. Other relevant factors

Personal and family characteristics

Additionally, the models in Table 3 include a large set of demographic and socioeconomic characteristics of the children. Here we report the most relevant associations between those characteristics and educational outcomes.

Estimates in Table 3 show that girls do significantly better than boys both in cognitive and in non-cognitive knowledge as measured by all four indicators. Conversely, Calero and Waisgrais (2008) found that girls had worse scores than boys in the 2006 PISA exams. The particular focus of PISA on Science, a subject in which boys tend to outperform girls, likely accounts for this difference. Interestingly, when we run the model for global knowledge separately by grade, the largest gender gap appears in sixth grade while the difference is very small for second and fourth grades. Across subjects, though, the gender gap reverses in second grade for maths scores, with boys outperforming girls, while girls outdo boys in written and oral expression as well as in reading throughout the three grades in the sample. Hence, there are clear differences between genders. It is important to point to some previous research that finds that certain teacher characteristics may play a role in explaining the difference in performance between boys and girls. Lavy (2008), for instance, shows that this is the case in Israel, where there is some favourable bias towards girls. Dee (2006) finds that learning from a teacher of the opposite gender has a detrimental effect on the academic progress of students and their engagement in school. The paper also shows that adverse gender effects have an impact on both boys and girls, but that the effect is greater in male students in middle school, simply because most middle-school teachers are women. This evidence implies that

(4) For the IV to work, the degree of involvement of grandparents would need to be unrelated to children's performance. The relation is not necessarily the same across countries. For instance, while we find that there is no relationship between the two variables in our dataset for Catalonia (within the limitations imposed by the information available), Loury (2006) finds some positive relation of grandparent involvement for the US.

controlling for teachers' gender would help to understand differences in academic outcomes between boys and girls. Unfortunately, we do not have any tools to check whether this is an important correlate in our sample since we do not know the gender of the teachers.

Regarding the gap in cognitive development across genders, it is worth mentioning the paper by Gneezy and Niederle (2003) that investigates whether men and women react differently in competitive environments. The authors show that men tend to do better than women in a competitive setting, which suggests that girls might benefit from further competitive stimulus at school in order to improve their performance.

With respect to the relationship between grade and children's development, the coefficients of the dummies for each grade are not statistically significant for the cognitive indicators and *School Abilities*. By contrast, ceteris paribus, students in fourth and, particularly, in sixth grade score worse in *Social Behaviour*. This is not surprising since they are closer to adolescence.

Results show that family structure matters since children who live in a *Rebuilt* family score significantly less in *Global Knowledge*, and *School Abilities* than children who live in a *Nuclear* family. Students from *Monoparental* and *Extensive* families also obtain lower marks than those in *Nuclear* families, although the coefficients are generally not statistically significant. Overall, this suggests that family stability tends to boost children's development. Of course, the data is too limited to allow for any causal inference on this.

In rich countries, the trade-off between the quantity and the quality of children partly accounts for the decrease in fertility. In this framework, families value not only the number of children, but also human capital investment in each child and, as a result, limit their offspring [Becker and Lewis (1973), Leibowitz (1974)]. In developing countries, where children contribute to the family income, a negative association between family size and educational outcomes would be expected [see Li *et al*. (2007) for China]. We find a negligible negative relationship between the number of siblings and all four development indicators in Catalonia [see Angrist *et al*. (2005) for a similar result in Israel].

Our dataset does not have complete information on child spacing but we know whether the child has siblings and whether he/she ranks first, middle or last. Powell and Steelman (1993) show that close sibling spacing increases the likelihood of dropping out of high school and decreases the odds of attending post-secondary school. Birth order has also been associated with differential earnings and educational attainment [Behrman and Taubman (1986), Hanushek and Kimko (2000), Black *et al*. (1991) and Booth and Kee (2009)]. Results in Table 3 do not display significant differences in *Global Knowledge* by birth order. Elder children and, especially, single children have significantly better Catalan skills than the rest, but they score low in non-cognitive knowledge, especially in *Social Behaviour*. Most likely, the firstborn tend to receive more adult attention early in life than their siblings and this boosts their language acquisition. Likewise, long hours of playing alone and not needing to share family resources with siblings may curtail the development of their social abilities.

First generation immigrants, that is, children born out of Spain, tend to have worse academic performance than those born locally, except for Asians [see Calero and Waisgrais (2008) for similar findings]. Interestingly, there are no performance differences between students born in Catalonia and those born in the rest of Spain. The gap in non-cognitive knowledge (that is, in *School Abilities* and *Social Behaviour*) mostly disappears for all immigrant groups. In separate regressions, we have studied whether there is any difference in the performance of children born in Catalonia of immigrant parents (second generation immigrants) and the other native children, and found no significant disadvantage for the former group. This indicates the presence of some assimilation of immigrant children. The coefficient for Catalan spoken at home, the main language of instruction in elementary schools, is positive but not significant. Language spoken at home, once origin is controlled for, is not a key factor for educational outcomes.

Table 3 shows a positive association between the monthly net income of the household and children's development. However, most of the coefficients are not significant once information on parents' education is included. In parallel regressions, we also added several combinations of parental education and the results showed some gains from pooling higher levels of education (i.e. two highly educated parents). Parental labour force status is not significant either. This suggests that the educational profile of the family matters more for children's development than the level of family income or the labour force status (see Hanushek and Kimko (2000) for a similar result). From another perspective, Ciccone and García-Fuentes (2008) also look at the gains in academic results from higher levels of parental education. More precisely, they estimate what would have been the average results of the PISA exams in certain regions if the average parental education in those areas was similar to the average education in the total sample of the country. They find some gains, although not very large, which suggests that there are other factors, such as the education system, that are also important.

School characteristics

Other things being equal, children in public schools receive higher scores in cognitive knowledge (i.e. *General Knowledge*, *Catalan Knowledge* and *School Abilities*) than those in private schools with partial public funding (*"concertada"*). Conversely, Calero and Waisgrais (2008) found no difference in learning outcomes from PISA surveys between the two types of schools. Although we do not have a clear prior, different grading practices among public and private school teachers (i.e. grade inflation in public schools) could account for this finding and for the disparity between the two studies. Given that school grades partly determine university access, this finding confirms the need to use more objective evaluations (i.e. standardised tests) at least to supplement school grades.

Time use by the student

Results in Table 3 suggest that studying languages, computing and music, as well as reading, enhance children's development. Sports activities improve educational achievement to a lesser extent. Hence, encouraging the taste for reading and participating in certain extra-curricular activities should be beneficial for children's development.

4. Conclusions

This paper investigates the relationship between socioeconomic characteristics and children's cognitive and non-cognitive development in Catalonia using data from the project "Família i Educació a Catalunya" of the Fundació Jaume Bofill collected in 2005. Cognitive knowledge is calculated with two quantitative indicators reported by the teacher and non-cognitive knowledge with information reported by both the teachers and the parents.

The main finding of the paper is that the quarter of the year when a child is born is associated with distinctive outcomes. The youngest students in a class tend to underperform the rest academically and this maturity gap does not decrease as the children advance into later grades. This suggests that decisions on cutoff rates are very important since they affect future learning outcomes. In this light, distributing children into grades with shorter age spreads, especially for early years, seems a reasonable policy. However, given the likely high costs of targeting groups of such small birth-intervals, alternative means to compensate for this gap, such as remedial tutoring for students born later in the year, seem appropriate.

Furthermore, our study finds some differences in children's development according to the age when they first enrolled in any formal school. Children who attended nursery school generally do better than those who started at the first year of preschool (*P3* at age three) or later. This association, however, disappears once place of birth and family education variables are included in the analysis Even though there are no large differences between those who start before the age of one, and the rest, results suggest that sometime between ages 1 and 2 is probably the optimal entry age.

Overall, these results confirm what previous literature has already noted in other contexts such as the US [Heckman and Masterov (2007), Heckman *et al*. (2006)]. Governments may want to allocate more funds to preschools and encourage earlier enrolment of children of first generation immigrants since this would clearly be beneficial for their later development. Whether these results are transferable to other contexts such as, for example, Nordic countries where formal education starts later in life but students tend to perform well in comparative surveys such as PISA, remains to be seen. Nonetheless, enrolment in nursery schools is widely extended in those countries.

Construction of cognitive and non-cognitive indicators

Global Knowledge: This is the arithmetic average of the valuations given by the teacher on seven subjects (Science, Catalan, Spanish, Foreign language, Mathematics, Art and Physical education) on the scale of either low (1), average (2) or high (3). The measure is missing for 58 out of 942 individuals, mostly because information is not available on a few of the seven subjects (the data is missing in all subjects only for 3 individuals). To check whether the missing observations are random, and do not bias the results, we undertake the following two analyses. First, we substitute the missing values in *Global Knowledge* by an eighth score (*Global Achievement*) that the teacher provides and that

roughly summarises the global performance of the child. In fact, the two measures, our *Global Knowledge* and *Global Achievement* are quite similar, with an average difference of only around 0.02. Substituting the missing observations in *Global Knowledge* by *Global Achievement* reduces the number of missing values to 19. Second, we use the sample bias correction model of Heckman (1979) to analyse whether the subsample containing missing observations on *Global Knowledge* is random. Although, a priori, there is no reason for the teacher to consistently overlook the score for a specific type of student, it is important to double check it and correct it with the Heckman adjustment if needed. Results, however, indicate that no correction is granted. In the main estimates presented in the paper, we use the first correction for missing values.

Catalan Knowledge: This is the arithmetic average of the following skills: writing, reading, oral expression and comprehension. The teacher assesses the knowledge from none (0) to very high (10). There are only 5 missing values.

School Abilities: This indicator has been built using the responses to questions on social competence and antisocial behaviour of children from the *School Social Behaviour Scales* (SSBS-2) in Merrell (2002). Only 23 observations are missing.

Social Behaviour: This indicator is built using the responses of the Strengths and Difficulties Questionnaire (SDQ) in Goodman (1997). Five dimensions of social behaviour are measured: the *emotional symptoms* score, the *conduct problems* score, the hyperactivity score, the *peer problems score* and the *prosocial behaviour* score. The first four scores are negatively oriented and can be regrouped into an indicator labelled *total difficulties* score. Goodman (1997) explains in *http://www.sdqinfo.com* how to build these scores from the questionnaire. The algorithm transforms responses from 25 variables (untrue, something is true and definitely true) plus a set of other variables that calibrate the impact of these factors on the wellbeing of the child into final scores (for each respondent). Our *Social Behaviour* is calculated as the average of the *prosocial behaviour score* for parents and teachers. Only 15 cases are missing. In social sciences, it is common practice to derive indices for competence and social behaviour from a mix of questionnaires. In this particular project, surveys with various sections (e.g. SSBS-2, SDQ, APQ, etc.) were addressed to parents, teachers and 12 year olds, and each contained. How to summarise all this information into a simplified indicator is not straightforward. Social researchers tend to use factor analysis to describe variability among observed variables (responses) in terms of fewer unobserved variables called factors. The observed variables are modelled as linear combinations of the factors, plus "error" terms. The information gained about the interdependencies of different responses can be used later to reduce the set of variables in a dataset. Using this statistical instrument, Bonillo *et al*. (2007) transform all the information of the surveys into two indicators (personal competence and pro-social-normative behaviour). We take a simplified version of this since we observe that, in fact, our four dependent variables are in line with the results of their factor analysis and the gains from undertaking factor analysis are small. Our cognitive measure and our non-cognitive measure *School Abilities*, for example, belong to their personal competence group, while the fourth indicator, the non-cognitive measure *Social Behaviour*, accounts for their pro-social conduct.

REFERENCES

Amarelo, C., Bové, N. (2008): "Situacions Familiars i Laborals: una Anàlisi de l'Enquesta de Població Activa des de la Perspectiva de les Llars", in Economia de les llars. Revista d'economia catalana i de sector públic, vol. 91, Departament d'Economia i Finances. Generalitat de Catalunya, pp. 105-118.

Angrist, J., Krueger, A. (1991): "Does Compulsory School Attendance Affect Education and Earnings?", *The Quarterly Journal of Economics*, vol. 106, n. 4, pp. 979-1014.

Angrist, J., Lavy, V., Schlosser, A. (2005): "New Evidence on the Causal Link Between the Quantity and Quality of Children", NBER WP No. 11835.

Beavers, W.R., Hampson, R.B., Hulgus, Y.F. (1990): "Beavers Systems Model Manual", Dallas, Texas, Southwest Family Institute.

Becker, G.S., Lewis, H.G. (1973): "On the Interaction Between the Quantity and the Quality of Children", *Journal of Political Economy,* vol. 81, n. 2, pp. 279-288.

Bedard, K., Dhuey, E. (2006): "The Persistence of Early Childhood Maturity: International Evidence of Long-Run Effects", *Quarterly Journal of Economics*, vol. 121, n. 4, pp. 1437-72.

Behrman, J., Taubman, P. (1986): "Birth Order, Schooling, and Earnings", *Journal of Labor Economics,* vol. 4, n. 3, pp. 121-145.

Black, S.E., Devereux, P.J., Salvanes, K.G. (1991): "The More the Merrier? The Effect of Family Size and Birth Order on Children's Education", *The Quarterly Journal of Economics,* vol. 120, n. 2, pp. 669-700.

Bonillo, A., Doval, E., Molinuevo, B., Pardo, Y., Pérez, C., Torrubia, R. (2007): "Família i Educació a Catalunya", Fundació Jaume Bofill.

Booth, A., Kee, H.J. (2009): "Birth Order Matters: the Effect of Family Size and Birth Order on Educational Attainment", *Journal of Population Economics,* vol. 22, pp. 367-397.

Bound, J., Jaeger, D. (2000): "Do Compulsory School Attendance Laws Alone Explain the Association between Quarter of Births and Earnings", *Research in Labor Economics,* vol. 19, n. 4, pp. 83-108.

Buckles, K., Hungerman, D. (2010): "Season of Birth and Later Outcomes: Old Questions, New Answers", Notre Dame and NBER, August.

Calero, J., Waisgrais, S. (2008): "Entorn Familiar i Rendiment Educatiu", in Economia de les llars. Revista d'economia catalana i de sector públic, vol. 91, Departament d'Economia i Finances, Generalitat de Catalunya, pp. 87-99.

Calero, J., Waisgrais, S. (2008): "Qué Determina el Rendimiento de los Alumnos Immigrantes? Una Primera Aproximación a Partir de PISA-06?", *Investigaciones de Economía de la Educación,* 3, pp. 153-162.

Cascio, E., Schanzenbach, D.W. (2007): "First in the Class? Age and the Education Production Function", NBER WP No. 13663.

Chay, K.Y., McEwan, P.J., Urquiola, M. (2005): "The Central Role of Noise in Evaluating Interventions that Use Test Scores to Rank Schools", *American Economic Review,* vol. 95, n. 4, pp. 1237-1258.

Ciccone, A., García-Fuentes, W. (2008): "The Quality of the Catalan and Spanish Education Systems: A Perspective from PISA", ICREA. UPF.

Cohen, J., Cohen, P., Aiken, L.S. (2003): "Applied Multiple Regression/Correlation Analysis for the Behavioral Sciences", 3rd edition, Mahwah, N.J.: Lawrence Erlbaum.

Crawford, C., Dearden, L., Meghir, C. (2007): "When You Are Born Matters: the Impact of Date of Birth on Educational Outcomes in England", Institute for Fiscal Studies.

Dee, T.S. (2006): "The Why Chromosome: How a Teacher's Gender Affects Boys and Girls", *Education Next*, vol. 6, n. 4, pp. 275-278.

Domingo, A., Bayona, J. (2008): "Migració Internacional i Llars a Catalunya" in Economia de les llars. Revista d'economia catalana i de sector públic, vol. 91, Departament d'Economia i Finances. Generalitat de Catalunya, pp. 51-69.

Dronkers, J. (1994): "The Changing Effects of Lone Families on the Educational Attainment of their Children in a European Welfare State", *Sociology,* vol. 28, pp. 171-192.

Elder, T., Lubotsky, D. (2009): "Kindergarten Entrance Age and Children's Achievement: Impacts of State Policies, Family Background, and Peers", *Journal of Human Resources,* vol. 44, n. 3, pp. 641-683.

Flaquer, L. (2008): "Diversitat Familiar, Benestar de la Infiıncia i Cohesió Social a Catalunya", in Economia de les llars. Revista d'economia catalana i de sector públic, vol. 91, Departament d'Economia i Finances. Generalitat de Catalunya, pp. 71-86.

Frick, P.J. (1991): "Alabama Parenting Questionnaire", University of Alabama.

Gneezy, U., Niederle, M. (2003): "Performance in Competitive Environments: Gender Differences", *The Quarterly Journal of Economics*, vol. 118, n. 3, pp. 1049-1074.

Goodman, R. (1997): "The Strengths and Difficulties Questionnaire: A Research Note", *Journal of Child Psychology and Psychiatry*, vol. 38, pp. 581-586.

Hanushek, E.A., Kimko, D.D. (2000): "Schooling, Labour Force Quality, and the Growth of Nations", *American Economic Review*, vol. 90, n. 5, pp. 1184-1208.

Heckman, J.J. (1997): "Sample Selection Bias as an Specification Error", *Econometrica*, vol. 47, n. 1, pp. 153-161.

Heckman, J.J., Masterov, D.V. (2007): "The Productivity Argument for Investing in Young Children", NBER Working Paper no. 13016.

Heckman, J.J., Stixrud, J., Urzua, S. (2006): "The Effects of Cognitive and Noncognitive Abilities on Labour Market Outcomes and Social Behaviour", *Journal of Labor Economics,* vol. 24, n. 3, pp. 411-482.

Jaffee, S., Caspi, A., Moffitt, T.E., Belsky, J., Silva, P. (2001): "Why are Children Born to Teen Mothers at Risk for Adverse Outcomes in Young Adulthood? Results from a 20-year Longitudinal Study", *Development and Psychopatology,* vol. 13, pp. 377-397.

Lavy, V. (2008): "Do gender stereotypes reduce girls' or boys' human capital out-comes? evidence from a natural experiment", *Journal of Public Economics,* vol. 92, n. 10-11, pp. 2083-2105.

Leibowitz, A. (1974): "Home Investment in Children", *Journal of Political Economy,* vol. 82, n. 2, pp. 111-131.

Li, H., Zhang, J., Zhu, Y. (2007): "The Quantity-Quality Tradeoff of Children in a Developing Country: Identification Using Chinese Twins", IZA DP No. 3012.

Loury, L.D. (2006): "All in the Extended Family: Effects of Grandparents, Aunts, and Uncles on Educational Attainment", *The American Economic Review*, vol. 96, n. 2, pp. 275-278.

McEwan, P.J., Shapiro, J.S. (2008): "The Benefits of Delayed Primary School Enrollment", *Journal of Human Resources*, vol. 43, n. 1, pp. 1-29.

Merrell, K.W. (2002): "School Social Behavior Scales", Paul H. Brookes Publishing Co.

Nelson, R.R., Phelps, E. (1996): "Investment in Humans, Technology Diffusion, and Economic Growth", *American Economic Review*, vol. 56, n. 2, pp. 69-75.

OECD (2003): "Where Immigrant Students Succeed: A Comparative Review of Performance and Engagement in PISA 2003", OECD.

OECD (2006): "Assessing Scientific, Reading and Mathematical Literacy: A Framework for PISA 2006", OECD.

Piancentini, J.C., Cohen, P., Cohen, C. (1992): "Combining Discrepant Diagnostic Information From Multiple Sources. Are Complex Algorithms Better Than Simple Ones", *Journal of Abnormal Child Psychology,* vol. 20, pp. 51-53.

Powell, B., Steelman, L.C. (1993): "The Educational Benefits of Being Spaced out: Sibship Density and Educational Progress", *American Sociological Review,* vol. 58, pp. 367-381.

Rebelo, S. (1991): "Long-Run Policy Analysis and Long-Run Growth", *Journal of Political Economy*, vol. 99, n. 3, pp. 500-521.

Romer, P. (1990): "Endogeneous Technological Change", *Journal of Political Economy*, vol. 99, n. 5, pp. 71-102.

Sánchez, A.P. (2008): "Efectos de la Immigración en el Sistema Educativo: el Caso Español", Tesis doctoral de la Universitat de Barcelona.

Stipek, D. (2002): "At What Age Should Children enter Kindergarten? A Question for Policy Makers and Parents", Social Policy Report: Giving Child and Youth Development Knowledge Away.

Weil, E. (2007): "When Should a Kid Start Kindergarten?", *The New York Times Magazine*, June 3, 2007.

Fecha de recepción del original: enero, 2011
Versión final: octubre, 2011

RESUMEN

Este estudio analiza la relación entre los factores socioeconómicos y el rendimiento educativo de los alumnos de primaria en Cataluña. El análisis se centra principalmente en el papel que tienen el trimestre de nacimiento y la edad de escolarización sobre el aprendizaje. Los datos, recopilados aleatoriamente en una muestra de 191 colegios durante el año 2005, contienen información sobre alumnos de segundo, cuarto y sexto curso de primaria. Uno de los resultados del análisis es que los estudiantes más jóvenes de la clase, nacidos cerca de la fecha de corte del 31 de diciembre, obtienen peores resultados académicos que el resto. Esta desventaja subsiste a lo largo del ciclo escolar. Por el contrario, las aptitudes sociales entre alumnos nacidos en distintos meses del año son similares. Este estudio también muestra que los alumnos que fueron escolarizados más pronto, especialmente antes de los tres años, presentan un mejor rendimiento. De todos modos, este resultado es menos significativo en cuanto se incluyen indicadores socioeconómicos como el origen y la educación de los padres. Entre otros factores demográficos, un entorno menos aventajado, como el que supone crecer en familias no nucleares o con un nivel de estudios bajo, dificulta el aprendizaje. Asimismo, los inmigrantes de primera generación tienden a obtener resultados algo peores que los de los alumnos nacidos en España, y la participación en ciertas actividades extra-escolares tiene beneficios claros para el rendimiento escolar.

Palabras clave: rendimiento escolar, trimestre de nacimiento, edad de escolarización.

Clasificación JEL: I20, J24.