The Impact of Family Size on School Achievement: Test Scores and Subjective Assessments by Teachers and Parents

Deborah A. Cobb-Clark[1,2] & Julie Moschion[1]

[1]Melbourne Institute of Applied Economic and Social Research
University of Melbourne
Vic 3010, Australia
d.cobb-clark@unimelb.edu.au
moschion@unimelb.edu.au
Tel: +61(0)3 8344 2100

[2]Institute for the Study of Labor (IZA)

Preliminary version – Please do not cite or put online

August 2013

JEL: J13, I21, I24
Keywords: family size, school achievement, education, Australia

Acknowledgements
The authors are grateful for financial support from an Australian Research Council Discovery Grant (DP110103456). This paper uses unit record data from Growing Up in Australia: the Longitudinal Study of Australian Children, conducted in partnership between the Australian Government Department of Families, Housing, Community Services and Indigenous Affairs (FaHCSIA), the Australian Institute of Family Studies (AIFS) and the Australian Bureau of Statistics (ABS). The findings and views reported in this paper are those of the authors and should not be attributed to FaHCSIA, AIFS, the ABS or the ARC.

Abstract

This paper presents estimates of the impact of family size on pupils' school achievement using data on national test scores across multiple subjects as well as subjective assessments of pupils' performance provided by teachers and parents, for a sample of primary school pupils aged 8/9 and 10/11 years old in the Australian state of Victoria. The impact of family size is identified via instrumental variables (IV), exploiting plausibly random differences in the gender mix of siblings and twin births. In contrast to much of the literature, our results provide strong evidence for the existence of a quantity-quality trade-off, with IV estimates suggesting an economically meaningful negative impact of having more than one sibling. In reading, for example, having more than one sibling leads to a decrease in achievement of about 29% of a standard deviation when assessed by test scores, 33% when assessed by teachers and 14% when assessed by parents. For numeracy, we find a decrease in outcomes of respectively 39%, 51% and 21% of a standard deviation.

1. Introduction

More than fifty years ago, Becker (1960) applied the principles of standard consumption theory to analyze the decision to have children within an economic framework. He and his co-authors subsequently conceptualized parents' fertility decisions as a trade-off in the number (quantity) of children that they choose to have versus the per-child investments (quality) that they choose to make in them (Becker and Lewis 1973; Becker and Tomes 1976). Since then, the relationship between family size and children's human capital has become one of the most frequently estimated relationships in social sciences with a vast literature concluding that children in larger families tend to have smaller human capital endowments (see Hanushek (1992); Schultz (2005) for a review). This negative relationship is likely to be spurious, however, if researchers do not account for the simultaneity in parents' choice of child quantity and child quality. Economists have turned to the plausibly exogenous increase in family size generated by multiple-births (Rosenzweig and Wolpin 1980; Black et al. 2010; 2010; Cáceres-Delpiano 2006; Angrist et al. 2010) or the gender mix of siblings (Goux and Maurin 2005; Conley and Glauber 2006; Angrist et al. 2010) in an effort to identify the causal effect of family size on child outcomes. In general, these studies generate very little evidence that increasing family size results in poorer outcomes for children (see Conley and Glauber 2006; Angrist et al. 2010), though there are important exceptions. Cáceres-Delpiano (2006), for example, finds that twin births reduce first-born boys' chances of attending private school, while Black et al. (2010) find that twin births have a negative effect on the IQ of older boys.

The goal of this paper is to contribute to this literature by analyzing the impact of family size on several unique measures of education achievement for pupils attending primary school in Australia. We are particularly interested in the following research questions: Do children in larger families suffer from a disadvantage in education achievement, as measured by standardized test scores in years 3 and 5 of primary school? Do teachers overestimate or underestimate this disadvantage? Are parents aware of this disadvantage? In addressing these questions, we make several contributions. First, we focus on family-size effects on the school achievement of primary school pupils rather than on later – and arguably cruder measures of – educational outcomes (e.g. educational attainment, grade retention, private school attendance, etc.). This is of interest because difficulty in early school years may have consequences on later educational experiences, even if they do not affect school attainment per se. Second, we investigate whether teachers exhibit stereotypes in relation to family size either by overestimating or underestimating family-size effects.[1] This is important because if teachers have negative stereotypes, this can lead pupils to exert lower efforts and achieve lower outcomes in a ‘self-fulfilling prophecy’ (Arrow, 1973; Mechtenberg, 2006). Moreover, there is some evidence that the lower expectations of teachers’ for children from stigmatized social groups may increase children’s difficulties in school (see review by Jussim & Harber, 2005). Third, we estimate family-size effects in parental reports of children’s achievement. These are likely to differ from those inherent in objective achievement measures, e.g. standardized test scores, because parents may have difficulties in evaluating their child’s performance relative to his peers or because they suffer from a declaration bias such that they systematically overestimate their child’s success, particularly if he does not perform very well. These estimates are important because parents are an essential source of support for children and it is important that they are aware of children’s difficulties. Finally, we use a birth-cohort data-set to provide the first evidence on the quantity/quality trade-off for Australia.

We identify the effects of family size on our objective and subjective achievement measures using a combination of the gender mix of siblings and twin births as instruments in an instrumental variables (IV) model.[2] While some researchers use one or the other, combining the two instruments broadens the population used in the estimation and increases external validity. Moreover, both instruments are potentially subject to omitted variables biases and “a comparison of twins and sex-composition estimates therefore provides a specification check since the omitted variables biases associated with each type of instrument should differ” (Angrist et al. 2010). We interpret the fact our two instruments yield similar results as evidence for the internal validity of our estimates.

We find strong evidence for the existence of an objective disadvantage stemming from growing up in a larger family, as measured by National Test Scores (NAPLAN). Specifically, the IV estimates suggest a meaningful negative impact of having more than one sibling on tests scores of around 29% and 39% of a standard deviation, respectively in reading and numeracy. Results also suggest a negative bias in teachers’ achievement towards larger families such that: having more than one sibling leads to a decrease in assessment by teachers of about 33% and 51% of a standard deviation, respectively in reading and numeracy. Parents’ assessment of the family-size disadvantage are

[1] There is some evidence of teacher biases with respect to other attributes. Lavy (2004) finds a negative bias in teachers’ assessment for boys in Israel. Gibbons and Chevalier (2008) find differences between objective and subjective assessments by socioeconomic status while Burgess and Greaves (2009) find differences between White and ethnic minority students in the UK. Cornwell & al (2013) investigate the source of this gender bias in teachers’ assessment and find evidence that these are related to difference in non-cognitive skills between boys and girls.

[2] Some papers only use gender composition instruments (Cruces and Galiani, 2007; Hirvonen, 2008; Baez, 2008); some others use twin births only (Rosenzweig and Wolpin, 2000; Black et al., 2005). A few papers use both instruments separately and /or simultaneously (Angrist and Evans, 1998; Angrist & al., 2010; Black et al. 2010).

smaller than test score estimates (respectively 14% and 21% of a standard deviation), suggesting that they minimize the objective disadvantage and ignore the subjective one induced by teachers.

The rest of the paper is structured as follows. Section 2 briefly reviews the existing literature and sets out our empirical approach. Section 3 describes the data and sample used. Our main results are discussed in Section 4, while our conclusions and suggestions for future research can be found in Section 5.

2. Estimating the Effect of Family Size on Child Outcomes:[3]

Much of the literature which seeks to understand the effect of family size on child outcomes estimates reduced-form relationships of the following form:

$$Q_i = \alpha + \gamma N_i + X_i \beta + \varepsilon_i \quad (1)$$

where i indexes families, Q is a measure of per-child human capital, N is the number of children, X is a vector of controls, ε is the error term, and the remaining terms are parameters to be estimated. Given the above, $\hat{\gamma}$ reflects the causal impact of the number of children (N) on average per-child quality (Q) so long as $COV(N_i \varepsilon_i) = 0$.

Identification:

There are two primary threats to causality: (i) unobserved heterogeneity or (ii) endogeneity in family size. In particular, it may be the case that families differ in unobserved ways – say in preferences for children, socio-economic status, genetic factors, etc. – that are correlated with both the quantity and quality of children they desire (see Guo and VanWey 1999). This leads $\hat{\gamma}$ to suffer from omitted variable (selectivity) bias. Alternatively, the simultaneous nature of the quality-quantity trade-off implies that family size is itself a function of children's average human capital endowments resulting in simultaneity (endogeneity) bias in $\hat{\gamma}$.

Some researchers have used sibling fixed-effects models to difference out the effects of unobserved heterogeneity in any family characteristics that are correlated with both decisions to have and to invest in children (e.g. Guo and VanWey 1999). This approach is useful in reducing the potential for

[3] De Haan (2010); Aslund and Gronqvist (2010).

omitted variable bias, but it suffers from two problems. Specifically, it cannot account for (i) unobserved heterogeneity that varies across time or siblings; or (ii) fertility decisions which depend in part on the quality of previous children (Conley and Glauber 2006).

Others have exploited the exogenous variation that naturally occurs in human fertility – i.e. multiple births and the gender mix of siblings – as instrumental variables in an effort to identity the causal effect of family size. IV estimation has been extremely useful in reducing the potential for estimation bias, however, there remain some potential pitfalls. First, different instruments identify the causal effect of increased family size for different populations of families (Angrist and Imbens, 1995). Twin studies typically identify the effect of having additional siblings due to twinning in families who would not have had this additional child otherwise (see Black et al. 2005, 2010; Angrist and Zhang 2006). Studies utilizing the sex composition of siblings rely on the observation that families are more likely to have additional children if their older children are of the same gender. Studies which use a same-sex instrument typically generate parameter estimates that apply to the population of families who go on to have an additional child because their first two children were of the same gender but would not have done so otherwise (Conley and Glauber 2006).[4] If those instruments generate different estimates (for different sets of compliers), it is difficult to make comparisons across studies using different instruments or to generalize results to the population of families as a whole. Using both instruments improves the prospects for comparison and generalization by broadening the population our estimates are based on.

Second, there is debate in the literature about the extent to which twins and same-sex instruments meet the necessary requirements for consistent estimation of family-size effects. Some have suggested that twin births may not be truly random events and instead may relate to unobserved family background effects (see Angrist et al. 2010; Black et al. 2010), while having two children of the same sex may not be random if selective abortion is widely practiced (see Lee 2008). Also contentious, is the possibility that neither instrument meets the exclusion restriction. For example, there may be economies of scale – e.g. in shared clothing, bedrooms, activities, etc. -- associated with having children of the same sex which would free up resources for families to make additional human capital investments (Rosenzweig and Wolpin 2000; Angrist et al. 2010). In contrast, families with twins may divert resources away from older children in an effort to compensate for the fact that twins typically have lower birth weight and may have worse health or cognitive achievement as a

[4] See Angrist et al. (2010) for a detailed comparison of the effects identified by these two instruments.

result (Rosenzweig and Zhang 2009). The available empirical evidence is extremely modest but points towards these issues not being critical in developed countries.[5]

Given the above discussion, we follow what is becoming standard practice and compare estimated family-size effects generated using both a twin and a same-sex instrument. Each captures a different source of fertility variation and, thus, each results in an estimate of the local average treatment effect (LATE) for a different subset of the population (i.e. compliers). We can increase external validity by using both instruments and broadening the sub-population of interest. Moreover, both instruments may be affected by omitted variables bias. Twin births, for example, may vary with mothers' age, race, health, etc. and may result lower birth weights and less spacing between children. Gender mix, on the other hand, may result in economies of scale associated with the sharing of clothing, toys, or bedrooms when children are of the same sex. Angrist et al. (2010) argues that a comparison of estimates from both instruments is an important specification check. Finding similar results using these different identification strategies is evidence that these instruments are internally valid. Finally, we separate out the effect of family size from the effect of being a twin by following others in the literature and restricting the analysis to first-born children (Black et al., 2010).

Estimation Model:

We estimate the following two-stage linear model. In the second stage, children's school achievement (Y_i) is explained by family size (X_i) and other covariates (W_i) in the following equation:

$$Y_{ij} = \alpha'_{0j} W_i + \beta_{1j} X_i + \varepsilon_{ij} \qquad (2)$$

where j successively indexes test scores, teachers' assessment and parents' assessment. Moreover, x_i is an indicator variable which takes a value of 1 if the family has more than two children and 0 otherwise. The first-stage equation links the family-size variables to the instruments. Specifically,[6]

$$X_i = \pi'_0 W_i + \gamma_1 Z_i + \eta_i \qquad (3)$$

[5] Abrevaya (2009) finds no evidence of gender selection among White Americans. In Australia, the probability to have a third child is similar between families with two first boys or two first girls suggesting no preference for boys in Australia (Moschion, 2013). Also, Butikofer (2011) test for the presence of economies of scale in consumption in families with children of the same sex in a variety of countries and finds no significant differences between the estimated equivalence scales of families with different sibling sex composition in richer countries.

[6] When the endogenous explanatory variable (fertility) is a dummy, another solution to endogeneity is the use of simultaneous equations with a probit regression in the first-stage (Heckman, 1978). Linear probability models impose no assumptions regarding the residuals and are also suitable to use when the instrument, the endogenous variable and the dependent variable are dummy variables (Heckman and Macurdy, 1985).

where Zi is a vector of instruments for having more than two children. “Twins 2” is an indicator variable equal to 1 if the mother had twins at her second birth (and 0 otherwise), while “same sex” is an indicator variable equal to 1 if the two eldest siblings are of the same gender (and 0 otherwise). The coefficient $\hat{\gamma}_1$ gives the effect of the instrumental variables on fertility.

Our contribution lies in analysing family-size effects on school achievement of primary school pupils and comparing those with family-size effects on teachers’ and parents’ subjective assessments. Also, we are able to combine twins and gender mix instruments in a specification that controls for some of the potential sources for omitted variable bias.

3. Data

The Longitudinal Study of Australian Children (LSAC) is an Australian national study designed to provide an in-depth understanding of children's development.[7] The study commenced in 2004 with the recruitment of two cohorts: one cohort of 5,107 children aged 0-1 year old (the birth or "B cohort") and another of 4,983 children aged 4-5 years old (the kindergarten or "K cohort") and their families across all states and territories of Australia. Interviews have since been conducted with families every two years.

School Achievement Measures

Of particular interest to us are the subjective evaluations of children's academic performance that are asked to the children’s teacher and a responding parent. In more than 96 percent of cases, the responding parent is the child’s birth mother.[8] In particular, teachers are asked “Overall how would you rate this child's academic skills, compared to other children of the same grade level?” Parallel questions are also asked about language and literacy skills and mathematical skills, respectively. Possible responses include: (i) far below average; (ii) below average; (iii) average; (iv) above average; and (v) far above average. Similarly, mothers are asked: “Compared to other children in their class how well do you think (study child) is progressing in reading (respectively maths)?” Mothers can respond by saying: (i) much better; (ii) a little better; (iii) about the same; (iv) a little worse; or (v) much worse. Mothers are also asked: “How would you describe (study child’s) overall achievement at school?” and can respond: (i) excellent; (ii) above average; (iii) average; (iv) below

[7] The study is a partnership between the Australian Government Department of Families, Housing, Community Services and Indigenous Affairs (FaHCSIA), the Australian Institute of Family Studies (AIFS) and the Australian Bureau of Statistics (ABS).

[8] For simplicity, we will refer to responding parents as “mothers”.

average; or (v) well below average. These assessments are available for children aged 8-9 (which corresponds to wave 3 for the K cohort) and 10-11 (wave 4 for the K cohort).[9] In constructing our measures, we have reversed the scaling of questions directed to mothers to make it consistent with teachers' assessments and our standardized test scores, so that higher scores reflect greater achievement.

LSAC data are linked to standardized test scores from the National Assessment Program - Literacy and Numeracy (NAPLAN) which assesses all Australian students in Years 3, 5, 7 and 9 in reading, writing, language conventions (spelling, grammar and punctuation) and numeracy, using a national test, administered on the same day, that has been conducted annually since 2008. The reporting scales from 0-500 are constructed so that given scale scores can be compared across school grades and over time. For example, a score of 500 in Reading for Year 3 in 2008 means the same as a score of 500 for Year 5 in 2009. The linkage from NAPLAN to LSAC was successful for 84 percent of the K cohort children. In order to draw comparisons with our subjective achievement measures, we focus mainly on reading and numeracy.[10]

Currently, NAPLAN test scores are available for Years 3 and 5 of the K cohort.[11] Because of different age regulations across States and parents' option to delay their children's school entry, children born the same year (in the same LSAC cohort-wave) may be enrolled in one of three different sequential grades. This poses difficulty in establishing a correspondence between NAPLAN scores and information collected in LSAC. Moreover, the timing of LSAC interviews (from March to December every two years) differs from that of NAPLAN (in May each year). Consequently, to minimize the time difference between our subjective and objective measures of children's achievement and to avoid using test scores from different grades for children in the same cohort-wave, we analyse Year 3 test scores for children aged 8-9 (cohort K, wave 3) and Year 5 test scores for children aged 10-11 (cohort K, wave 4). Some of these children will have taken the NAPLAN test before the interview, others after. For example for NAPLAN Year 5, LSAC Wave 4: 23 percent of children took the NAPLAN test in 2009, one year before the LSAC interviews; 71 percent took the NAPLAN test on the same year as the interview (2010) (87 percent of whom took the test before the interview); and 6 percent took it one year later (2011).

[9] Wave 5 for the B cohort, containing both subjective assessment and NAPLAN test scores will be available soon and added to the analysis.
[10] However, the analysis has been conducted on the other subjects as well and yield similar results.
[11] Unfortunately, 23 percent of Year 3 test scores are missing because children were enrolled in Year 3 in 2007, before the NAPLAN tests were introduced.

Controls

We use the detail of the LSAC data to control for the mother's characteristics (six dummies for education, sex, age, age at first birth, a dummy for whether she is born in Australia, the age she arrived in Australia, a dummy for indigenous status, 5 dummies for religion, 8 state dummies, a dummy for the language first spoken not being English, a dummy for not speaking English well), children's characteristics (sex of the study child, log of birth weight of the study child, age of second or third child) and the wave (which characterises the age of the study child). There are a number of valuable additions to usual controls. First, we control more extensively for the mothers' origins by adding controls for the age of arrival in Australia, religion and English proficiency. Second, we are able to control for the birth weight of the study child and in robustness checks for the mother's BMI. This is of importance because, at least in some countries, mothers giving birth to twins have been shown to have higher BMIs (Bhalotra and Clarke, 2013) and thus potentially also bigger babies (outside of the twin births). Finally, we control for the age of the second child (in families with two children) or the third child (in families with at least three children). This allows us to control for the fact that third children from a twin second birth are on average older than those from singleton births. Following the literature, we analyse the outcomes of first-born children to deal with both birth order effects and twins specificities which may affect our outcomes.

Estimation Sample

Our sample is constituted of children from cohort K, Waves 3 and 4 (aged 8-9 and 10-11) for whom we have complete NAPLAN test scores and complete subjective assessments from their teacher and responding parent. Focusing on primary school is important for two reasons. First of all, at high-school level, parents might be less aware of their children's academic achievements and, secondly, high school students have a number of different teachers and, therefore, any one teacher cannot properly assess a child's entire academic performance. Our sample is then restricted to first-born children who live with two biological or adoptive parents and only have biological or adoptive siblings in families with at least two children (around 1,815 children).[12]

[12] Although, our sample size is not huge, our first stages are very similar to other results in the literature and we are able to reproduce very close estimates of the impact of family size on mothers' labour market participation to Moschion (2013). Specifically, we find that having more than two children decreases mothers' labour market participation by 23 percentage points when using gender mix as an IV (sd: 0.13), by 16 percentage points when using twins as an IV (sd: 0.05) and by 17 percentage points when using both IVs simultaneously (sd: 0.05) (comparable results from Moschion (2013) are respectively: 0.24 (0.07), 0.11 (0.03) and 0.12 (0.03)). Our results are not dramatically less precise and may be a bit larger because the LSAC children are younger than the Census children.

4. Instruments validity

Excludability of the Instruments

Instrumental variable estimates converge to the true parameter value as long as the instrument is correlated with the outcome only through the endogenous explanatory variable, i.e. $COV(N_i \varepsilon_i) = 0$ in equation (1). Specifically, this implies that gender mix and twin births affect children's outcomes only because it increases family size. This implies that the instrument is random: no pre-treatment difference in factors related to children's outcomes exists between those receiving the treatment (i.e. having twins or same-sex siblings) and those not receiving it. Basically, this can be simply formulated as: does every family have the same probability of having same-sex eldest siblings (twins)? There is no direct test of this assumption but it has been widely tested indirectly in the literature by comparing the average characteristics of those affected by the instrument and those unaffected. Table 1 gives summary statistics for children in our sample and documents differences in these characteristics depending on whether children are from a family with same-sex children or from a family with different-sex children (respectively twins versus non-twins).

Our results with respect to gender mix are similar those in the previous literature. In particular, we find that gender mix does not depend on family background characteristics such as the parent's age, age at first birth, level of education, being born in Australia, indigenous status, state of residence and the first-born's sex (see Table 1). Interestingly, it does not vary either when we look at more detailed information provided by LSAC on the parent's age of arrival in Australia, English proficiency and the study child's birth weight. There is also no evidence of imbalance with respect to the wave of data. We do however find that the second/third child is about 3 months younger in families with same-sex siblings (which is to be expected since third children are on average younger than second children and happen more often in families with same-sex siblings) and that the proportion of individuals with no religion is lower in these families by about 6 percentage point (while families with different sex siblings have a higher probability of being Anglican). When adding all these characteristics as controls in first stage regressions of family size on gender mix, the coefficient of gender mix does not vary significantly (Table 2).[13]

The story with respect to twin births is more complex. It has been well documented that the probability to have twins increases with age and fertility treatments so that twin births are found to be correlated with certain characteristics such as age for example. In our sample, we find that mothers of twins are older and had their first child later (see Table 1). We also find that parents with twins have slightly more education, are less often indigenous, are disproportionately from a non-Christian religion, have lower English ability and arrive in Australia younger, all of which may simply reflect different ages at maternity. Parents of twins are less likely

[13] For example, for the 'NAPLAN test scores' sample, the difference between controlling and not controlling for characteristics is 0.097–0.052 = 0.045 with a standard error of $(0.027^2 + 0.023^2)^{0.5}$ = 0.036.

to live in the Northern Territory or the Australian Capital Territory, but the number of observations in these states is quite small such that no twins are observed there. As expected, we also find that families with twins have an older third/second child than other families, such that controlling for this variable will capture the effect of the specificity that twins have the same age. Black et alii. (2010) additionally provide some evidence that the birth weight of the first child is slightly correlated with twins at second birth, suggesting that healthier women more often give birth to both heavier babies and twins. As a result, without controlling for the weight of the first child, the impact of family size (instrumented by twins) may capture as well a positive "healthy child effect" and thus bias our estimates upward. In our sample, twins' older sibling is actually 300 grams lighter than non-twins older sibling.

All in all, our instruments do not show hugely unexpected correlations and controlling for the wide array of characteristics available, we assume that our instruments are orthogonal to remaining unobservable factors that could affect the outcome.

Strength of the Instruments

The consistency of our instrumental variable estimates also relies on the fact that family size and the instruments are sufficiently correlated (Bound, Jaeger and Baker, 1995). If the correlation between the instruments and family size is weak, even a small correlation between the instrument and the error term can bias the estimates more than the OLS. A common rule of thumb is that in first-stage regressions with no controls, the F-statistic should be larger than 10. Table 2 provides first-stage results, i.e. the effect of gender mix and twins at second birth on family size. The F-statistics in models without controls (first, third and fifth column) are at least equal to 10, which suggest that our instruments are strong enough to produce consistent instrumental variable estimates.

In models with controls, we find that coming from a family with same-sex siblings increases the probability of having more than one sibling by 5.2 percentage points compared to families with different-sex siblings. And having twins as younger siblings increases the probability of having more than one sibling by 84.1 percentage points. Interestingly, these effects remain similar when both instruments are used together (5.4 and 84.3 respectively) suggesting that they capture different LATEs, making it sensible to use the two instruments together.

5. Results

NAPLAN Test Scores

Results for family-size effects on NAPLAN Test scores are reported in Table 3.[14] OLS estimates (first column) suggest practically no association between family size and children's test scores: first-born children who have two siblings or more do not have lower test scores than those with one sibling only, apart from slightly higher numeracy scores and a slightly lower probability to score in the bottom two quintiles in numeracy (these effects are not significant). Adding controls to OLS regressions (second column) makes these positive associations between family size and test scores even more positive by controlling for characteristics which are knowingly positively associated with family size and negatively with children's test scores (as parental education for example). This suggests that children from larger families do not suffer more difficulties in tests than their counterparts from smaller families. A possible reason for these non-negative OLS estimates is that higher quality births entails parents to have an additional child, creating a fallacious positive correlation between family size and the first-born outcomes. Also, it is possible that the first-born benefits from an additional younger sibling by interacting with him or teaching him.

Turning to IV estimates, the impact of family size appears to be clearly negative in both numeracy and reading, although coefficients are not always significant. The addition of a second sibling to first-born children decreases reading test scores by 26 points and numeracy test scores by 32 points out of 500 (eighth column). This is equivalent to a respective decrease of 29 percent and 39 percent of a standard deviation in test scores.

2SLS estimates are larger than their OLS counterparts, suggesting an upward bias in OLS estimates which is mostly at odds with the literature. This would imply that Australian families with more than two children have some unobserved characteristic (or behavior) that leads their children's test scores to be higher in comparison to those of children in two-child families. Moreover, these characteristics are orthogonal to our instruments -- parents with same-sex siblings or twins do not differ in this respect from other parents -- such that 2SLS estimates corrects for this bias. This could cast some doubt on our 2SLS estimates but other elements are quite reassuring. First, first-stage results and results for mothers' labour market participation are consistent in magnitude with other results in the

[14] For ease of comparison with teachers' and parents' assessments, we report only results for Reading and Numeracy but the analysis has been performed for Spelling, Grammar and Writing and yield similar results. Results are available upon request.

literature, This is particularly true in regards to the comparison between OLS and 2SLS: estimates using gender mix are slightly larger than the OLS and those with twins slightly smaller as in Moschion (2013). Also, comparing results in the fourth and sixth column, our 2SLS estimates using one of the two instrument only are close to each other suggesting that an omitted variable bias (specific to the instrument) is unlikely and that the family-size effect on test scores is probably quite homogenous among the two different populations of compliers (or that these compliers are representative of the same population). Finally, some recent papers also find positive biases in OLS estimates of the effect of family size on children's outcomes. For example, Mogstad and Wiswall (2012) find small OLS estimates of the effect of family size on children's outcomes which they interpret as misleading. After correcting for endogeneity and allowing for a non-monotonic relationship between family size and children's education, they find strong evidence of negative family-size effects.

Finally, we construct indicator variables for our outcomes which characterize more precisely the full distribution of results and are mirror the measures used to capture teachers' and parents' subjective assessments of children's achievement. Specifically, we construct an indicator which equals 1 if the child's score is in the bottom (respectively top) two quintiles of the distribution of scores in the subject for this cohort and wave. Consistent with results at the mean, having more than one sibling increases the probability of being in the bottom of the distribution and decreases the probability of being at the top. More precisely, an increase in family size particularly increases the probability of being at the bottom of the reading distribution and decreases the probability of being at the top of the numeracy distribution.

Teachers' Assessment of Child Achievement

Table 4 presents results for family-size effects on whether the teacher thinks that the study child is far below average, below average, average, above average or far above average, in reading, mathematics and overall. Just as for NAPLAN test scores, OLS estimates are small and insignificant with negative coefficients on family-size decreasing when controls are introduced. Interestingly however, once standardized, it appears that teachers overestimate (relatively to NAPLAN test scores) the educational disadvantage associated with family size. Specifically, children with more than one sibling receive teachers' assessment that are lower by 7 percent of a standard deviation in reading (coefficient of -0.060) and by 5 percent in mathematics (coefficient of -0.038) compared to NAPLAN test scores which are lower by 1 percent and higher by 7 percent respectively. However, the differences in test-based and teacher-based assessments of student performance are small and could be due the bias associated with the endogeneity of family size.

2SLS estimates are again more negative than their OLS counterparts. And interestingly, the difference between family-size effects objectively assessed by test scores and that subjectively assessed by teachers is maintained once the endogeneity of family-size is accounted for. Having more than one sibling decreases the teacher's assessment by 0.285 points in reading and 0.416 in numeracy out of a 5-points scale. In standardised terms, this amounts to respectively 33 percent and 51 percent of a standard deviation. This suggests that teachers either overestimate the difficulties of pupils from larger families (stereotype) or perceive some difficulty not revealed by test scores (possibly related to non-cognitive skills as suggested by Cornwell et al. 2013). The difference between objective and subjective assessments in the 2SLS also suggests that this additional difficulty identified by teachers is actually a consequence of family size (and not other things correlated with family size).

Here, there does not appear to be specific non-linearities in achievement as all quintiles are affected by family size. Specifically, being from a family with more than two children increases the probability of being in the bottom two quintiles and decreases the probability of being in the top two quintiles both in reading and numeracy.

Parents' Assessment of Child Achievement

Table 5 provides the results for parents' assessment of whether their child is much better, a little better, about the same, a little worse or much worse than his peers. Results suggest smaller and non-significant family-size effects compared to those for test scores. First, OLS are mainly small and insignificant. Second, 2SLS suggest that having more than one sibling has a negative impact on children's achievement, but this effect is smaller in magnitude to those estimated on test scores: the coefficients of -0.146 and -0.215 for reading and numeracy amount to decreases of 14 percent and 21 percent of a standard deviation in parents' assessment. This suggests that parent perceive some of the disadvantage of their children but only partly. The analysis of non-linearities are consistent with small negative effects of family-size according to parents, which, if anything, come from higher probabilities to rank in the bottom.

Robustness checks

Our sample of responding parents include both mothers and fathers. When, we restrict our sample to mothers, we find that they also ignore family-size effects. Controlling for the age the child was when taking the test yields very similar results and controlling for the mother's BMI tends to reduce the sample and the effects slightly (effects for numeracy remain significant) but we still find that teachers tend to overestimate family-size effects while parents underestimate them.

Another potential source of concern about our instruments are potentials for post-treatment differences (i.e. the fact that the effect of the instrument on the outcome may go through another channel than family size) and direct effects of the instrument on the outcomes. Here the concern is especially salient as we found 2SLS effects which are larger than OLS estimates. Again, there exists no direct test for these potential issues but our evidence seems to suggest that post-treatment differences and direct effects do not bias our 2SLS estimates. Over-identification tests fail to reject the excludability of our instruments (Hansen P-values are all above 30% for NAPLAN test scores; above 19 percent for teachers' assessments and above 5 percent for parents' assessments). Although Hansen P-values are not extremely high for teachers' and parents' assessments in reading (19 percent and 8 percent respectively), we find that 2SLS estimates of family-size effects remain unchanged when using one instrument only and controlling for the other instrument (to allow for direct effects of the instrument on the outcome). Consistent with the high Hansen P-values, 2SLS estimates using gender mix and twin births separately are always consistent in sign and mostly quite close and not significantly different (apart from those generated using parents' assessment in reading). Although our instruments are quite different and subject to different type of biases, they yield consistent estimates of family size-effects suggesting again that potential biases should not be affecting our 2SLS estimates significantly. This also suggests that the two different populations of compliers experience similar family-size effects providing some evidence that our estimates are not totally specific to a very small subset of the population.

6. Conclusions and Outlook

We find that increasing family-size from two to more than two children results in an objective disadvantage in children's achievement through test scores and an additional subjective disadvantage expressed by teachers (either because of stereotypes or because teachers have a more detailed view of children achievements). Parents do not seem to be fully aware of this disadvantage.

Results for wave 3 age 8-9 are slightly lower than for 10-11 but this could come from the fact that this cohort is censored because NAPLAN did not exist when 20 percent of this cohort was in Yr3 (in 2007). The wave 5 of data, available in September, will provide data for the B cohort at ages 8-9, thus providing additional sample as well as a way to test whether this censoring affects the results.

We will investigate further whether the difference in objective and subjective assessment by teachers could be explained because teachers perceive better children's ability not captured in test scores. In particular we will investigate the role of non-cognitive skills (Cornwell & al, 2013).

References:

Abravaya, J. (2009), ‘Are There Missing Girls in the United States? Evidence from Birth Data’, *American Economic Journal: Applied Economics*, 1(2), 1-34.

Angrist, J. D., and Imbens, G. W. (1995), “Average causal response with variable treatment intensity”, *Journal of the American Statistical Association*, 90, 431–42.

Angrist, J. D. and Evans. W. N. (1998), “Children and Their Parents’ Labor Supply: Evidence from Exogenous Variation in Family Size”, *American Economic Review*, 88(3), 450–77.

Angrist, J. D., Lavy, V. and Schlosser, A. (2010), “Multiple experiments for the causal link between the quantity and quality of children”, *Journal of Labor Economics*, 28, 773-824.

Arrow, K. (1973). The Theory of Discrimination. *In Orley Ashenfelter and Albert Rees, eds. Discrimination in Labor Markets. Princeton, N.J.: Princeton University Press*, 3-33.

Baez, J. E. (2008), “Does More Mean Better? Sibling Sex Composition and the Link between Family Size and Children’s Quality”, *IZA Discussion Paper*, 3472.

Bhalotra, A. and Clarke, D. (2013), “Child Quantity versus Quality: Are Twin Births Exogenous?”, mimeo

Becker, G. S. (1960) "An Economic Analysis of Fertility," in Demographic and economic change in developed countries. NBER Special Conference Series No. 11. Princeton: Princeton University Press, 1960, pp. 209-31.

Becker, G. S., and Lewis, H. G. (1973), ‘On the interaction between the quantity and quality of children’, *Journal of Political Economy* 81: S279–S288.

Becker, G. S., and Tomes, N. (1976), ‘Child endowments and the quantity and quality of children’, *Journal of Political Economy* 84:S143– S162.

Black, S. E., Devereux, P. J. and Salvanes, K. G. (2005), “The More the Merrier? The Effect of Family Size and Birth Order on Children’s Education”, *Quarterly Journal of Economics*, 120(2), 669-700.

Black, S. E., Devereux, P. J. and Salvanes, K. G. (2010), “Small Family, Smart Family? Family Size and the IQ Scores of Young Men”, *Journal of Human Resources*, 45(1), 33-58.

Burgess, Simon, and Ellen Greaves. 2009. “Test scores, Subjective Assessments and Stereotyping of Ethnic Minorities” Working Paper

Butikofer, A. (2011), ‘Sibling Sex Composition and Cost of Children’, Mimeo

Caceres-Delpiano, Julio. (2006), ‘The impacts of family size on investment in child quality’, *Journal of Human Resources* 41:722–37.

Conley, D. and Glauber, R. (2006), “Parental Educational Investment and Children’s Academic Risk: Estimates of the Impact of Sibship Size and Birth Order from Exogenous Variation in Fertility”, *Journal of Human Resources*, 41(4), 722–737.

Cornwell, Christopher M., David B. Mustard, and Jessica Van Parys (2013), “Non-Cognitive Skills and the Gender Disparities in Test Scores and Teacher Assessment – Evidence From Primary School” *Journal of Human Resources* 48(1), 236-264.

De Haan, M. (2010), ‘Birth order, family size and educational attainment’, *Economics of Education Review*, 29(4), 576-588.

Gibbons, S., & Chevalier, A. (2008), ‘Assessment and age 16+ education participation’, *Research Papers in Education, 23*(2), 113-123.

Goux, D. and Maurin, E. (2005), "The effect of overcrowded housing on children's performance at school", *Journal of Public Economics*, 89(5-6), 797-819.

Hanushek, Eric A. (1992), ‘The trade-off between child quantity and quality’, *Journal of Political Economy* 100:84–117.

Heckman, J.J. (1978), ‘Dummy Endogenous Variables in a Simultaneous Equation System’, *Econometrica* 46(4), 931–959.

Jussim, L., & Harber, K. D. (2005), ‘Teacher expectations and self-fulfilling prophecies: Knowns and unknowns, resolved and unresolved controversies’, *Personality and Social Psychology Review, 9*(2), 131-155.

Lavy, V. (2008), “Do Gender Stereotypes Reduce Girls’ or Boys’ Human Capital Outcomes? Evidence from a Natural Experiment”, *Journal of Public Economics* 92(10): 2083–105.

Mechtenberg, L. (2006), ‘Cheap Talk in the Classroom: How biased grading at school explains gender differences in achievements, career choices, and wages’, *Forthcoming; Review of Economic Studies*.

Magne, M., and Wiswall, M. (2012) “Testing the Quantity-Quality Model of Fertility: Linearity, Marginal Effects, and Total Effects”, Working Paper

Rosenzweig, M. R., and Wolpin, K. I. (1980), ‘Testing the quantityquality fertility model: The use of twins as a natural experiment’, *Econometrica* 48:227–40.

Rosenzweig, M. R., and Zhang, J. (2009), ‘Do population control policies induce more human capital investment? Twins, birth weight, and China’s “one child” policy’, *Review of Economic Studies* 76:1149– 74.

Schultz, T. Paul (2005), ‘Effects of fertility decline on family well-being: Evaluation of population programs’, Draft for MacArthur Foundation Consultation meeting, Philadelphia

Table 1: Descriptive Statistics by Gender Mix and Twins 2

Variable	All	SS	Not SS	Difference	T2	Not T2	Difference
Wave 4	54.88	54.43	55.28	-0.01	58.97	54.79	0.04
	(49.78)	(49.83)	(49.75)	(0.01)	(49.83)	(49.78)	(0.06)
A. Characteristics of the respondent parent							
Education: year 12 or equiv.	11.40	11.54	11.29	0.25	7.69	11.49	-3.79
	(31.80)	(31.97)	(31.66)	(2.11)	(27.00)	(31.89)	(5.46)
Certificate	25.90	25.17	26.54	-1.37	10.26	26.24	-15.98***
	(43.82)	(43.43)	(44.18)	(2.60)	(30.74)	(44.01)	(5.66)
Advanced diploma/dipl.	10.36	10.84	9.93	0.91	25.64	10.02	15.62*
	(30.48)	(31.11)	(29.92)	(1.72)	(44.24)	(30.04)	(8.64)
Bachelor degree	22.42	22.38	22.47	-0.09	23.08	22.41	0.67
	(41.72)	(41.70)	(41.76)	(2.62)	(42.68)	(41.71)	(7.59)
Graduate diploma/certif.	8.48	8.74	8.25	0.49	0.00	8.67	-8.67***
	(27.87)	(28.26)	(27.53)	(1.81)	(0.00)	(28.15)	(0.94)
Postgraduate degree	9.09	8.39	9.72	-1.33	7.69	9.12	-1.43
	(28.76)	(27.74)	(29.64)	(1.77)	(27.00)	(28.80)	(4.28)
Sex	3.20	3.26	3.13	0.13	5.13	3.15	1.98
	(17.59)	(17.78)	(17.43)	(0.97)	(22.35)	(17.48)	(4.45)
Age	38.73	38.71	38.75	-0.04	40.44	38.70	1.74**
	(4.59)	(4.47)	(4.68)	(0.24)	(4.23)	(4.59)	(0.80)
Age at first birth	28.80	28.79	28.81	-0.02	30.44	28.76	1.67**
	(4.49)	(4.38)	(4.60)	(0.23)	(4.26)	(4.49)	(0.84)
Not born in Australia	22.09	20.63	23.41	-2.78	20.51	22.13	-1.62
	(41.50)	(40.49)	(42.36)	(3.17)	(40.91)	(41.52)	(7.65)
Age arrived in Australia	3.93	3.69	4.14	-0.45	2.51	3.96	-1.44*
	(8.87)	(8.64)	(9.07)	(0.67)	(5.98)	(8.93)	(0.82)
Indigenous	1.38	0.93	1.78	-0.84	0.00	1.41	-1.41***
	(11.66)	(9.62)	(13.22)	(0.63)	(0.00)	(11.78)	(0.41)
Religion: None	22.96	19.72	25.86	-6.14*	30.77	22.79	7.98
	(42.07)	(39.81)	(43.81)	(3.22)	(46.76)	(41.96)	(11.07)
Catholic	27.98	26.95	28.90	-1.95	41.03	27.69	13.33

	(44.90)	(44.40)	(45.35)	(2.81)	(49.83)	(44.76)	(10.35)
Anglican	22.52	25.44	19.90	5.54**	17.95	22.62	-4.67
	(41.78)	(43.58)	(39.94)	(2.62)	(38.88)	(41.85)	(8.10)
Other christian	20.47	21.59	19.48	2.11	10.26	20.70	-10.44*
	(40.36)	(41.17)	(39.62)	(2.79)	(30.74)	(40.53)	(6.26)
Other religion	6.07	6.30	5.86	0.44	0.00	6.20	-6.20***
	(23.89)	(24.31)	(23.51)	(1.68)	(0.00)	(24.13)	(0.97)
English is a 2nd language	15.21	13.64	16.61	-2.98	20.51	15.09	5.42
	(35.92)	(34.34)	(37.24)	(2.25)	(40.91)	(35.81)	(7.22)
Does not speak English well	1.38	1.17	1.57	-0.40	0.00	1.41	-1.41***
	(11.66)	(10.74)	(12.43)	(0.89)	(0.00)	(11.78)	(0.39)
B. Characteristics of the study child							
Sex	52.18	53.03	51.41	1.62	38.46	52.48	-14.02
	(49.97)	(49.94)	(50.01)	(3.33)	(49.29)	(49.95)	(11.21)
Birth weight (in grms)	3,340	3,309	3,367	-58	3,038	3,346	-308
	(624)	(633)	(614)	(39)	(965)	(613)	(207)
C. Characteristics of the family							
State: NSW	33.00	32.75	33.23	-0.48	46.15	32.71	13.44
	(47.04)	(46.96)	(47.13)	(3.15)	(50.50)	(46.93)	(10.37)
VIC	26.17	26.46	25.91	0.54	23.08	26.24	-3.16
	(43.97)	(44.14)	(43.84)	(2.18)	(42.68)	(44.01)	(8.38)
QLD	14.88	15.27	14.52	0.74	12.82	14.92	-2.10
	(35.60)	(35.99)	(35.25)	(1.79)	(33.87)	(35.64)	(6.34)
SA	7.38	7.34	7.42	-0.08	5.13	7.43	-2.30
	(26.16)	(26.10)	(26.22)	(1.39)	(22.35)	(26.24)	(4.10)
WA	11.68	11.54	11.81	-0.27	7.69	11.77	-4.08
	(32.13)	(31.97)	(32.29)	(1.19)	(27.00)	(32.23)	(6.67)
TAS	3.86	3.73	3.97	-0.24	5.13	3.83	1.30
	(19.26)	(18.96)	(19.54)	(1.15)	(22.35)	(19.19)	(4.56)
NT	0.55	0.70	0.42	0.28	0.00	0.56	-0.56**
	(7.40)	(8.34)	(6.45)	(0.33)	(0.00)	(7.48)	(0.29)
ACT	2.48	2.21	2.72	-0.50	0.00	2.53	-3**
	(15.55)	(14.72)	(16.27)	(0.62)	(0.00)	(15.72)	(1.04)
Age of the 2nd/3rd child	5.75	5.61	5.88	-0.27*	7.31	5.72	1.59***

	(2.21)	(2.22)	(2.19)	(0.14)	(1.67)	(2.21)	(0.32)
N	1,815	858	957	1,815	39	1,776	1,815

Levels of significance: *10%; **5%; ***1%.
Sample: first-born child in families with at least two biological or children (and all outcomes available).
Data: LSAC, cohort K, Waves 3 and 4.

Table 2: The Impact of Gender Mix and Twins 2 on Family Size

		Outcome: More than 2 children					
		OLS Same sex	OLS Same sex	OLS Twins 2	OLS Twins 2	OLS SS & T2	OLS SS & T2
NAPLAN test scores	Same sex	0.097***	0.052**	-	-	0.099***	0.054**
	se	(0.027)	(0.023)	-	-	(0.029)	(0.023)
	Twins 2	-	-	0.657***	0.841***	0.661***	0.843***
	se	-	-	(0.016)	(0.042)	(0.019)	(0.044)
	F	13	5	1,633	395	634	195
	N	1,815	1,802	1,815	1,802	1,815	1,802
Controls:		N	Y	N	Y	N	Y

Levels of significance: *10%; **5%; ***1%.
Sample: first-born child in families with at least two biological or children (and all outcomes available).
Note: controls include an indicator for the wave, characteristics of the respondent parent (6 education dummies for year 12 or equivalent, Certificate, Advanced diploma/diploma, Bachelor degree, Graduate diploma/certificate, Postgraduate degree; sex; age; age at first birth; a dummy for not being born in Australia; the age the parent arrived in Australia; a dummy for being indigenous; 5 religion dummies for no religion, catholic, anglican, other christian, other religion; a dummy for speaking English as a second language; a dummy for whether the parent does not speak English well); characteristics of the study child (gender, birth weight in log); family characteristics (8 state dummies; age of the second child in two-children families or age of the third child in families with more than two children). Standard errors are robust and clustered at the individual level.
Data: LSAC, cohort K, Waves 3 and 4.

Table 3: The Impact of Family Size on Children's Test Scores

		Outcome: NAPLAN test scores							
		OLS	OLS	2SLS Same sex	2SLS Same sex	2SLS Twins 2	2SLS Twins 2	2SLS SS & T2	2SLS SS & T2
Reading	More than 2	-0.716	0.646	-32.989	-17.701	-33.542	-26.477	-33.428*	-26.079
	se	(4.120)	(5.016)	(35.422)	(62.484)	(22.406)	(17.300)	(18.451)	(16.675)
	N	1,815	1,802	1,815	1,802	1,815	1,802	1,815	1,802
Numeracy	More than 2	5.568	7.778*	-36.959	-30.779	-44.021*	-32.580*	-42.564*	-32.498*
	se	(3.618)	(4.445)	(44.189)	(74.420)	(24.958)	(17.864)	(21.822)	(17.527)
	N	1,815	1,802	1,815	1,802	1,815	1,802	1,815	1,802
Reading in bottom	More than 2	0.029	0.017	0.326	0.489	0.263*	0.184*	0.276**	0.197*
two quintiles	se	(0.022)	(0.030)	(0.236)	(0.475)	(0.146)	(0.105)	(0.127)	(0.107)
	N	1,815	1,802	1,815	1,802	1,815	1,802	1,815	1,802
Reading in top	More than 2	0.005	0.034	-0.264	-0.344	-0.164	-0.118	-0.185*	-0.129
two quintiles	se	(0.022)	(0.029)	(0.251)	(0.445)	(0.113)	(0.088)	(0.099)	(0.085)
	N	1815	1802	1815	1802	1815	1802	1815	1802
Numeracy in bottom	More than 2	-0.024	-0.052*	0.311	0.420	0.166	0.090	0.196	0.105
two quintiles	se	(0.027)	(0.031)	(0.261)	(0.513)	(0.154)	(0.106)	(0.130)	(0.103)
	N	1815	1802	1815	1802	1815	1802	1815	1802
Numeracy in top	More than 2	0.020	0.044	-0.420	-0.687	-0.225*	-0.152	-0.266**	-0.177*
two quintiles	se	(0.028)	(0.034)	(0.321)	(0.617)	(0.120)	(0.100)	(0.117)	(0.100)
	N	1815	1802	1815	1802	1815	1802	1815	1802
Controls:		N	Y	N	Y	N	Y	N	Y

Levels of significance: *10%; **5%; ***1%.
Sample: first-born child in families with at least two biological or children (and all outcomes available).
Note: controls include an indicator for the wave, characteristics of the respondent parent (6 education dummies for year 12 or equivalent, Certificate, Advanced diploma/diploma, Bachelor degree, Graduate diploma/certificate, Postgraduate degree; sex; age; age at first birth; a dummy for not being born in Australia; the age the parent arrived in Australia; a dummy for being indigenous; 5 religion dummies for no religion, catholic, anglican, other christian, other religion; a dummy for speaking English as a second language; a dummy for whether the parent does not speak English well); characteristics of the study child (gender, birth weight in log); family characteristics (8 state dummies; age of the second child in two-children families or age of the third child in families with more than two children). Standard errors are robust and clustered at the individual level.
Data: LSAC, cohort K, Waves 3 and 4.

Table 4: The Impact of Family Size on Teachers' assessment of Children Achievement

		Outcome: Teachers' assessment							
				2SLS	2SLS	2SLS	2SLS	2SLS	2SLS
		OLS	OLS	Same sex	Same sex	Twins 2	Twins 2	SS & T2	SS & T2
Reading	More than 2	-0.060	-0.014	-0.931*	-1.338	-0.346	-0.235	-0.466**	-0.285**
	se	(0.043)	(0.057)	(0.495)	(0.908)	(0.232)	(0.155)	(0.187)	(0.142)
	N	1,815	1,802	1,815	1,802	1,815	1,802	1,815	1,802
Numeracy	More than 2	-0.038	0.009	-0.420	-0.622	-0.651**	-0.406*	-0.604***	-0.416**
	se	(0.043)	(0.053)	(0.438)	(0.803)	(0.280)	(0.223)	(0.230)	(0.208)
	N	1815	1802	1815	1802	1815	1802	1815	1802
Overall	More than 2	-0.035	0.011	-0.525	-0.722	-0.377*	-0.256	-0.407**	-0.277*
	se	(0.042)	(0.047)	(0.468)	(0.799)	(0.220)	(0.166)	(0.188)	(0.158)
	N	1815	1802	1815	1802	1815	1802	1815	1802
Reading in bottom	More than 2	0.009	0.010	0.197	0.222	0.161	0.096	0.168*	0.101
two quintiles	se	(0.018)	(0.021)	(0.183)	(0.361)	(0.110)	(0.072)	(0.094)	(0.068)
	N	1,815	1,802	1,815	1,802	1,815	1,802	1,815	1,802
Reading in top	More than 2	-0.031	-0.009	-0.518	-0.803	-0.104	-0.073	-0.190*	-0.106
two quintiles	se	(0.022)	(0.033)	(0.321)	(0.556)	(0.133)	(0.090)	(0.108)	(0.084)
	N	1815	1802	1815	1802	1815	1802	1815	1802
Numeracy in bottom	More than 2	0.028	0.023	0.192	0.289	0.222*	0.137	0.216**	0.144*
two quintiles	se	(0.018)	(0.021)	(0.189)	(0.377)	(0.122)	(0.090)	(0.098)	(0.082)
	N	1815	1802	1815	1802	1815	1802	1815	1802
Numeracy in top	More than 2	0.003	0.035	-0.159	-0.281	-0.272***	-0.155*	-0.248***	-0.161*
two quintiles	se	(0.026)	(0.031)	(0.269)	(0.474)	(0.095)	(0.091)	(0.095)	(0.088)
	N	1815	1802	1815	1802	1815	1802	1815	1802
Overall in bottom	More than 2	0.009	0.012	0.065	0.053	0.096	0.050	0.089	0.050
two quintiles	se	(0.018)	(0.020)	(0.153)	(0.322)	(0.124)	(0.090)	(0.109)	(0.088)
	N	1815	1802	1815	1802	1815	1802	1815	1802
Overall in top	More than 2	-0.031	-0.003	-0.435	-0.674	-0.221	-0.146	-0.265**	-0.170*
two quintiles	se	(0.026)	(0.030)	(0.376)	(0.606)	(0.136)	(0.100)	(0.127)	(0.097)
	N	1815	1802	1815	1802	1815	1802	1815	1802
Controls:		N	Y	N	Y	N	Y	N	Y

See notes to Table 3.

Table 5: The Impact of Family Size on Parents' assessment of Children Achievement

		Outcome: Parents' assessment							
		OLS	OLS	2SLS Same sex	2SLS Same sex	2SLS Twins 2	2SLS Twins 2	2SLS SS & T2	2SLS SS & T2
Reading	More than 2	-0.007	0.059	-1.229**	-1.789	-0.132	-0.068	-0.358	-0.146
	se	(0.068)	(0.072)	(0.583)	(1.222)	(0.270)	(0.190)	(0.218)	(0.176)
	N	1,815	1,802	1,815	1,802	1,815	1,802	1,815	1,802
Numeracy	More than 2	0.034	0.120	-0.287	-0.159	-0.403	-0.217	-0.379	-0.215
	se	(0.059)	(0.083)	(0.561)	(0.985)	(0.340)	(0.261)	(0.287)	(0.250)
	N	1815	1802	1815	1802	1815	1802	1815	1802
Overall	More than 2	0.025	0.125*	-0.295	-0.138	0.031	0.058	-0.036	0.049
	se	(0.044)	(0.067)	(0.447)	(0.821)	(0.237)	(0.156)	(0.209)	(0.150)
	N	1815	1802	1815	1802	1815	1802	1815	1802
Reading in bottom	More than 2	0.028*	0.018	0.287**	0.361	0.097	0.076	0.136	0.089
two quintiles	se	(0.015)	(0.016)	(0.142)	(0.251)	(0.099)	(0.069)	(0.085)	(0.067)
	N	1,815	1,802	1,815	1,802	1,815	1,802	1,815	1,802
Reading in top	More than 2	0.011	0.042	-0.482*	-0.759	0.002	0.035	-0.098	-0.001
two quintiles	se	(0.031)	(0.034)	(0.276)	(0.567)	(0.105)	(0.078)	(0.091)	(0.074)
	N	1815	1802	1815	1802	1815	1802	1815	1802
Numeracy in bottom	More than 2	0.021	0.010	0.342**	0.524	0.100	0.060	0.150	0.081
two quintiles	se	(0.015)	(0.020)	(0.168)	(0.335)	(0.112)	(0.084)	(0.093)	(0.079)
	N	1815	1802	1815	1802	1815	1802	1815	1802
Numeracy in top	More than 2	0.038	0.063	-0.161	-0.174	-0.121	-0.053	-0.130	-0.059
two quintiles	se	(0.025)	(0.039)	(0.277)	(0.477)	(0.152)	(0.115)	(0.134)	(0.112)
	N	1815	1802	1815	1802	1815	1802	1815	1802
Overall in bottom	More than 2	0.005	0.007	0.062	0.044	0.090	0.072	0.084	0.071
two quintiles	se	(0.011)	(0.016)	(0.094)	(0.166)	(0.104)	(0.073)	(0.088)	(0.070)
	N	1815	1802	1815	1802	1815	1802	1815	1802
Overall in top	More than 2	0.023	0.074**	-0.164	-0.164	0.003	0.037	-0.031	0.028
two quintiles	se	(0.025)	(0.033)	(0.278)	(0.504)	(0.103)	(0.074)	(0.095)	(0.070)
	N	1815	1802	1815	1802	1815	1802	1815	1802
Controls:		N	Y	N	Y	N	Y	N	Y

See notes to Table 3.