CLASS SIZE

IN THE EARLY YEARS:

IS SMALLER REALLY BETTER?

May 2001

Maria Iacovou

Institute for Social and Economic Research

Essex University

Wivenhoe Park

Colchester CO4 3SQ

U.K.

Telephone: +44 (0) 1206 873994

E-mail: maria@essex.ac.uk

Abstract:

Other things being equal, theory would suggest that students in smaller classes at school should do better in terms of attainment; convincing experimental evidence for this also exists in the US. However, a relationship between small classes and better outcomes has not generally been evident in individual-level studies, possibly because of endogeneity arising from low-attaining or otherwise ‘difficult’ students being put into smaller classes than their higher-achieving counterparts. This paper uses data from the National Child Development Study to estimate the effects of class size. Ordinary Least Squares estimates indicate that small classes are not related to attainment; however, Instrumental Variables estimates, with class size instrumented by the interaction between school size and school type, show a significant and sizeable association between smaller classes and higher attainment in reading in the early years of school. This effect is common to different groups of students, and for some groups (girls, and those from larger families), this association is also found to persist through to age 11.

Acknowledgements:

Thanks are due to the ESRC and the Institute for Social and Economic Research for financial support in carrying out this work; and to the Data Archive at Essex University for access to the NCDS data. I would like to thank participants at UCL and ISER workshops and an ESPE conference session for useful comments and suggestions; thanks go also to Richard Blundell, Costas Meghir, Christian Dustmann, John Ermisch, Marco Francesconi and Amanda Gosling. All deficiencies and errors in the paper are of course my own.

1 INTRODUCTION

Class size is a contentious issue in Britain. Under successive Conservative governments between 1979 and 1997, class sizes in state schools increased steadily, growing to become among the largest in the OECD (OECD 1996; Blatchford and Mortimore, 1994). In 1995, OFSTED, the government's schools inspection agency, produced a report declaring that

> "[although] small class sizes are of benefit in the early years of primary education…. reducing class size across the board by even a small amount is expensive; there is no evidence to justify this investment." (OFSTED 1995).

Despite having been extensively criticised for political bias and unsound methodology (Brimblecombe 1994; Bennett 1994; Day et al. 1996), this report was used extensively by the government of the day to back up its claim that large class sizes in Britain had no adverse effects on pupils.

The Labour party made a major issue of infant class sizes in its 1997 election manifesto; reducing infant class sizes to a maximum of 30 was one of its most widely-touted pledges, both in its election manifesto and during its first term in office.

> "Research evidence shows the importance of class size for younger children… We have pledged to reduce class sizes for 5, 6 and 7 year-olds to 30 or below within the next five years…. There will be a cost to introducing smaller classes, which we intend to meet through phasing out the Assisted Places Scheme[1]." (DfEE 1997)

In terms of research evidence rather than politics, the evidence on class sizes is very mixed. A large number of studies have been undertaken, some finding a sizeable and significant relationship between class size and student outcomes; some finding no discernible relationship; and some finding a relationship in the 'wrong' direction (i.e. that larger classes are associated with better outcomes). In general, experimental studies tend to show that reducing class sizes is beneficial (Achilles 1996; Finn and Achilles

[1] The Assisted Places Scheme was a programme introduced by the Conservative government which paid school fees, in part or in full according to a means test, to enable children from less well-off backgrounds to attend private schools.

1990; and Word et al. 1990a and 1990b); and studies using aggregate-level data (eg, statewide average levels of educational spending in the US) also find a positive relationship between educational resources and outcomes (Card and Krueger 1992a, 1992b, 1998; and Loeb and Bound 1996). However, studies using individual-level data have generally failed to find a conclusive relationship between class size and educational outcomes. This probably arises because being in a small class is associated with other factors leading to poor outcomes, such as being in a bottom or remedial stream, or having poor previous educational attainment. In general, the only individual-level studies which *do* find a relationship between class size and student attainment in the expected direction are those which use some means to overcome problems of endogeneity of the class size variable (Angrist and Lavy 1999; Case and Deaton 1999; Akerheilm 1995).

Although several individual-level studies of class size exist in the UK, including those of Davie, Butler and Goldstein (1972); Robertson and Symons (1996); Dearden, Ferri and Meghir (1997); Dolton and Vignoles (1998), and Feinstein and Symons (1999), none of these has found a significant relationship between class size and student outcomes. This paper uses the same data as the above studies: the extraordinarily rich National Child Development Study (NCDS). It differs from earlier research in that (a) it deals specifically with class size (rather than with educational resources generally) as a determinant of outcomes; (b) it has as its main focus class sizes in the early years, the age group around which debate in this country is currently centred; and (c) it makes use of an exogenous instrument for class size not previously exploited: the interaction between school size and school type. Using this exogenous instrument for class size reveals a significant and sizeable positive association between small classes and higher reading scores.

2 DATA

This investigation is based on data from the National Child Development Study (NCDS). This is a longitudinal study which takes as its subjects all children born in the week of 3rd - 9th March 1958. It was originally conceived as a one-off perinatal mortality study, and the first wave of data, collected shortly after the subjects' birth, contains detailed medical and socioeconomic histories of their families.

TABLE 1:
INFORMATION AND SAMPLE SIZES FOR WAVES 0-3 OF NCDS

	Age	Number of Observations	Areas covered
Perinatal Mortality Survey	Birth	18553	Maternal health, parental characteristics, perinatal medical details
1st Follow-up (wave 1)	7	12619	Parental characteristics School characteristics Medical examination Test scores
2nd follow-up (wave 2)	11	12131	Parental characteristics School characteristics Medical examination Test scores
3rd follow-up (wave 3)	16	10538	Parental characteristics School characteristics Medical examination Test scores Individual interview

Note: sample sizes for Waves 1-3 refer to the number of observations where students are in mainstream state schools with 20 or more pupils, in classes between 20 and 45, and with non-missing data for family background and school characteristics.

Rollow-up studies were carried out when the cohort was aged 7, 11, 16, 23 and 33. The studies between ages 7 and 16 contain information on educational attainment, schooling, health, and family circumstances. Those at ages 23 and 33 give a detailed account of the subjects' health, labour market behaviour and family and household situation since age 16.

Apart from being relatively large in size, this data set has three important advantages for investigating the effect of class sizes on attainment. Firstly, because it is a longitudinal study, one is able to observe not only the instantaneous effect of class size on attainment, but also how persistent these effects are over time. Secondly, the data set provides a rich set of controls for the home environment and family background. Thirdly, and unusually for a data set so rich in individual-level data, there is also a limited amount of information available on the subject's school environment.

Table 2 shows the distribution of class sizes in the NCDS sample: they were rather larger than those in Britain today, with an average class size slightly over 35 in Waves 1 and 2, and with only around one in five children taught in classes of 30 or less. This was in part due to a severe shortage of teachers at the time caused by the unexpected baby boom of the late 1950s and early 1960s.

TABLE 2:
DISTRIBUTION OF CLASS SIZES IN THE NCDS.

		Wave 1	Wave 2
Range		2 – 72	1 – 90
Mean Class Size		35.9	35.2
% of classes 30 and under		17.9	20.5
"	31-35	21.3	24.4
"	36-40	38.4	36.4
"	Over 40	22.4	18.7

Note: Figures are for students in mainstream state schools (infant, junior and combined primary schools). Those in special schools, independent schools, or schools classified as 'other' have been dropped from the sample.

Some extremely small class sizes are reported, which most likely indicate the size of the year group rather than the number of pupils per teacher in very small rural schools. All students in classes smaller than 20 have therefore been excluded from the analysis. Students in very large classes (all those over 45) have also been dropped from the data set for further analysis, as have the small numbers of children attending independent schools; those with special educational needs who attended special schools; and children attending a type of school classified in the data as 'other'.

Throughout this paper, test scores are normalized to have a mean of zero and a standard deviation of one. Thus, all coefficients may be read in units of one standard deviation.

3 CLASS SIZE: ARGUMENTS AND EVIDENCE

> “Form sizes of 22 boys and teaching group sizes often smaller than that ensure the personal attention of staff”
>
> “The School can offer your daughter ... maximum individual attention in classes which are kept small at *all* levels.”
>
> “Small classes (usually a maximum of 16), a well qualified and enthusiastic staff .... bring out the best in each child.”

The above extracts, from a handbook carrying advertisements for fee-paying schools[2], give some anecdotal evidence that schools believe that small classes are a selling point, and that many parents are willing to pay substantial amounts for their children to be educated in small classes[3].

> “How many parents engage in home schooling so their child (children) will be in a large class? How many parents pay tuition at private schools so that their children will be in large classes?… Why are “important” classes… smaller than “average” classes in schools? Why are classes for students who need special help smaller than classes for “average” students?”

This second extract is taken from Achilles (1996), writing about the results of Tennessee’s Project STAR (Student Teacher Achievement Ratio). This project, implemented as part of a statewide drive to improve school standards, followed the progress of approximately 7000 children in a range of inner-city, urban, suburban and rural schools, from their kindergarten year in 1985 to the end of third grade in 1989, and for a further two years after the experiment formally ended. Pupils in participating schools were randomly allocated to one of three types of class: ‘small’ (13-17 pupils); ‘regular’ (22-25 pupils), or ‘regular’ with a full-time classroom assistant; participating schools were required to have at least one of each type of class. Analysis of the project was undertaken by an internal team of researchers: Achilles (1996), Finn and Achilles (1990), and Word et al. (1990a and 1990b), an additional follow-up study was

[2] Independent Schools Yearbook 1994-95

[3] and, in this case, with a certain peer group, and a certain curriculum.

conducted when the children were in secondary school (the Lasting Benefits study, Nye et al 1994); and the data were subsequently re-analysed by Krueger (1999).

The main findings of the project are as follows. On all measures of achievement, students performed significantly better in small classes, though no effect was found from having a teacher aide. The effect was largest in the case of minority students and those in inner-city areas, and was apparent in every year of the study, but was most pronounced at first grade level. The benefits of small classes in the early years also persisted until the children were in seventh grade (even though by seventh grade none of the children had been in a small class for four years). Krueger's (1999) re-analysis of the STAR data confirms the findings of the internal researchers, and suggests in addition that the effect from being in a small class is more or less confined to the *first* year of being in such a class, possibly as a result of some 'school socialisation' effect.

Project STAR is the best-known experimental study of class sizes, but by no means the only one. A similar study (Project Prime Time) was conducted in Indiana between 1984 and 1986; again, the project was favourably evaluated with a recommendation that it should continue (Mueller, Chase and Walden 1988; Achilles 1990).

Despite this experimental evidence, the debate about class sizes (and the effect of educational resources in general) is very much open. Experiments are costly to replicate, so most of the research in this field has been done using existing data sets.

As mentioned earlier, studies using aggregate-level data (for example, statewide averages of educational expenditures), of the type conducted by Card and Krueger (1992a, 1992b, 1998) and Loeb and Bound (1996), tend to find a positive relationship between educational resources (generally educational expenditure, rather than class sizes or pupil teacher ratios) and student outcomes.

However, critics have drawn attention to the probable existence of aggregation bias: by aggregating data to the level of the state, omitted variables which operate at that level (for example, state-specific determinants of performance such as state political variables) bias estimates more seriously than they would at other (lower) levels of aggregation (Hanushek, Rivkin and Taylor 1996)[4]. Grogger (1996) finds that aggregating measures of schooling inputs by state yields much higher estimates of the effect of school expenditure on earnings than aggregating measures of inputs by school district; and that when district-level measures are used, estimates of the effect of schooling inputs are positive but extremely small.

Studies using individual-level data in general yield much more mixed and inconclusive evidence on the effects of class sizes (or school resources) than do the studies using aggregate data. Hanushek (1996) surveys 377 studies in this area (including 277 estimates of the effect of pupil-teacher ratios on student performance), and concludes that there is no evidence that increasing the level of school inputs would have any effect on student performance. Blatchford and Mortimore (1994) provide a comprehensive survey of research pertaining to the UK, most of which indicates that *large* classes are associated with better student attainment. The studies of Morris (1959), CACE (1967), Davie et al (1972), and Little et al (1973), all point in this direction. Later UK studies which fail to find that small classes or increased resources have a beneficial effect include Dolton and Vignoles (1998) Dearden, Ferri and Meghir (1997), Robertson and Symons, (1996) and Feinstein and Symons (1999).

Why should these individual-level studies fail to find a link between class size and pupils' performance? One possibility is that there are benefits associated with smaller classes, but not in the range being studied, where class sizes are already 'small enough'.

Another possible explanation has to do with the method of instruction: class size may matter more under some teaching methods and less under others (Bennett 1998; Dolton and Vignoles 1998). Where teaching is 'child-centred', the effects of the dilution of teacher time in a large class are likely to be relatively strong; if lessons are taught in a 'lecture' format, on the other hand, class size may not matter so much[5]. OFSTED (1995) and Blatchford and Mortimore (1994) note that if the method of instruction is related to class size, this may bias estimates of the effect of class size: if teachers use 'child-

[4] Interestingly, in the UK aggregation bias would tend to operate in the opposite direction, depressing rather than exaggerating the effect of educational resources, since at local authority level, educational funding is allocated in part according to assessed needs.

[5] Even if whole-class teaching methods are used, the teacher's time will still be diluted by student numbers in tasks such as marking and student discipline, and so larger classes might still be expected to affect student attainment adversely, albeit to a lesser extent

centred' methods in small classes but 'whole-class' methods in large classes, and if whole-class teaching is more effective than child-centred teaching, then large classes may appear to be more effective than small classes.

However, evidence from the educational literature suggests that teachers do not in fact modify their teaching methods a great deal when faced with differing class sizes. Bennett (1998) mentions three studies where teaching methods were found to be unrelated to class size. For example, Shapson et al. (1980) randomly assigned teachers to classes of four sizes ranging from 16 to 37 pupils; observation of classroom processes revealed that teachers did not change their methods substantially when faced with different-sized classes.

A further explanation of why class size may appear unrelated to student outcomes is if students vote with their feet: better schools will be fuller and have larger classes than poorer schools. Good schools may also attract parents from more privileged socio-economic groups, or parents who care more about their children's education. Both these factors may give rise to a spurious relationship connecting large classes with good scholastic outcomes.

While this may indeed be a problem in a situation where parents are able to choose their children's schools, it is unlikely to be an issue for the NCDS cohort, as during the period covered by this survey, schools took their students from quite strictly defined catchment areas and the capacity of a school was determined by the local authority. Of course, certain types of parents may have deliberately relocated to within the catchment area of a 'good' school, but the NCDS contains sufficient data on parental characteristics to control for this.

Since teacher quality is not observed in the NCDS data[6], estimates of the effects of class size may be biased if teacher quality is related to class size. The direction of any relationship between teacher quality and class size is not obvious *a priori*: it may be that 'good' teachers have more market power and are able to negotiate smaller classes for themselves; on the other hand, if schools are concerned with equality of outcomes for students, better teachers may be allocated to larger classes. Dolton and Vignoles (1998)

[6] The NCDS contains no information on the characteristics of individual teachers, other than the sex of the teacher in the second follow-up

and Dearden, Ferri and Meghir (1998) use local authority data on average teachers' pay as an explanatory variable; both find this to be unrelated to educational outcomes, although this may be because in the UK aggregate teachers' pay is an extremely poor indicator of teacher quality[7].

Another source of bias, and the one on which this paper focuses, arises if class size is related to students' ability. If more able students are allocated to larger classes, class size will be endogenous in the educational production function, and the estimated effect of being in a small class will be biased downwards – perhaps to the extent that the effect appears to be negative. *All* the individual-level studies which have found small classes to be beneficial to students - those of Angrist & Lavy (1999), Akerheilm (1995) and Case and Deaton (1999) - have found some means of controlling for the relationship between school resources and background characteristics. The section which follows indicates that there is a strong relationship between ability and class size in the NCDS data, and explains an instrumental variables estimation strategy for obtaining unbiased estimates of the effects of class size.

[7] Teachers in the UK were (and still are) paid according to a formula with increments for each year of service up to a maximum, with higher scales and additional allowances for heads and teachers with extra responsibilities (CACE 1967). Teachers' pay is therefore to some extent related to experience, but not to any other measure of teacher quality.

4 ENDOGENOUS CLASS SIZE

Class size for the NCDS sample is related to background variables at two levels at least. The first is at the level of grant allocation to local councils and schools. During the period covered by the NCDS, schools were paid for by a combination of funds from local and central government, with money from central government being allocated via a mechanism which was to some extent redistributive (CACE 1967; Simon 1991).

There is also evidence from the NCDS that resources were allocated *within* schools on a redistributive basis: lower-attaining children were put into smaller classes. Figures 1, 2 and 3 show the distribution of class sizes at ages 7, 11 and 16 by ability groups, or streams[8]. At all ages, the distribution for top streams lies to the right of the distribution for middle streams, while the distribution for bottom streams lies well to the left.

Columns I, II and V of Table 3 display the results of OLS regressions with class size regressed on three dummies for ability group. At all ages students in top ability groups are in significantly larger classes than those in bottom groups, with a difference of around five pupils. In columns III and VI an additional indicator of ability (reading score at age 7) has been added into the regressions for ages 11 and 16, and in both regressions yields a positive coefficient significant at the 1% level. In other words, more able students find themselves in larger classes *even after* accounting for the effects of streaming. Thus, the problem of endogeneity cannot be overcome simply by controlling for a child's stream in regressions. Nor can it be overcome by restricting the sample to those in unstreamed classes: columns IV and VII in Table 3 show a relationship between class size and prior attainment significant at the 10% level at age 11 and the 1% level at age 16 in unstreamed classes. No indicator of prior ability is available for Wave 1, but the evidence from later sweeps suggests strongly that class size is endogenous at age 7.

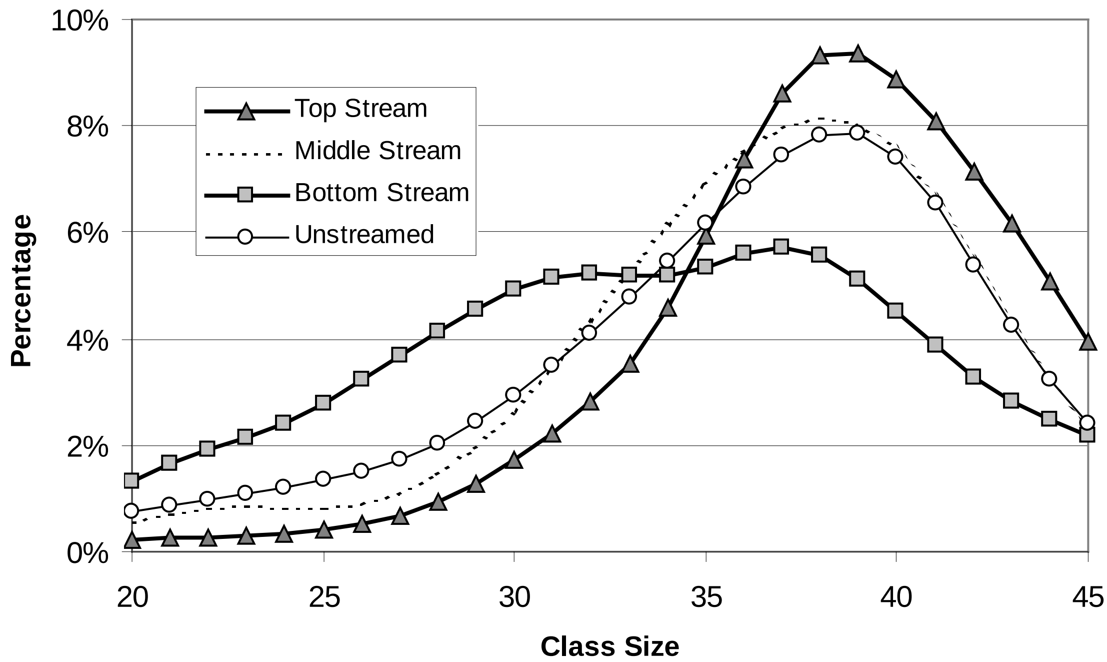
Possible explanations for this phenomenon are (1) class sizes are smaller in more difficult areas; (2) an effect operates at the margin, whereby particularly difficult pupils are assigned (say) to the smaller of two parallel classes; or (3) some classes which are coded in the data as unstreamed (for example, parallel classes) in the data set are in fact streamed[9].

---

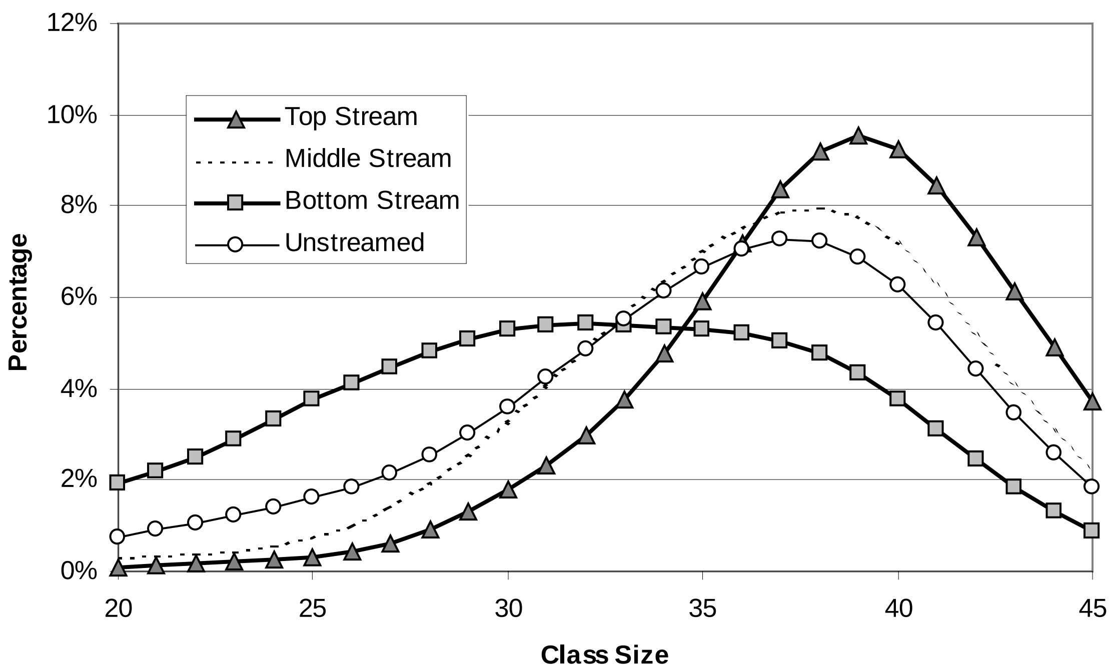
[8] .For readability, smoothed kernel densities are shown The Nadaraya-Watson kernel estimator has been used, with a smoothing bandwidth of 2.

[9] Jackson (1964), in a large survey of primary schools in 1962, found that by the age of seven almost three-quarters of children in schools large enough to have more than one class per year, were streamed. This is a far larger proportion than the 6 or 7 per cent reported in the NCDS in 1964. It does seem likely, therefore, that some of the classes coded in the data at Wave 1 as 'parallel' or 'divided by age within year groups' were in fact streamed.

Distribution of class sizes at age 7, by ability group



Distribution of class sizes at age 11, by ability groups



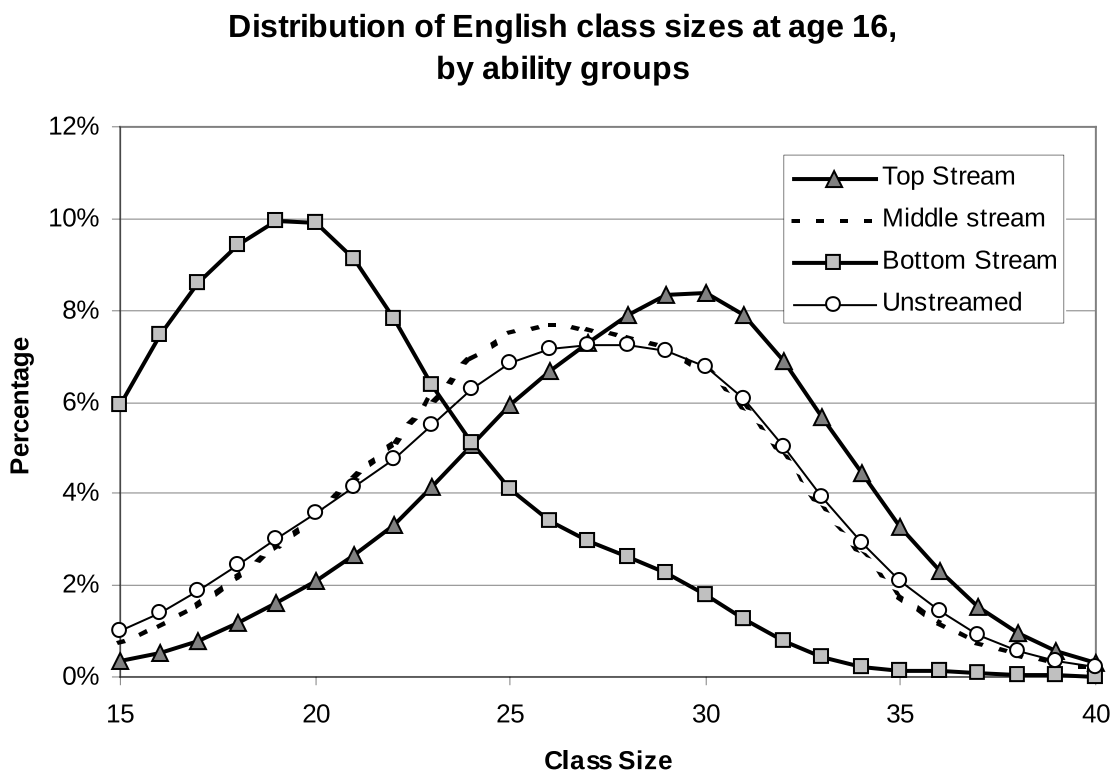

Distribution of English class sizes at age 16, by ability groups
Top Stream
Middle stream
Bottom Stream
Unstreamed
12%
10%
8%
6%
4%
2%
0%
Percentage
15
20
25
30
35
40
Class Size

TABLE 3:
OLS REGRESSIONS OF CLASS SIZE ON ABILITY GROUP AND PAST ATTAINMENT

	I	II	III	IV	V	VI	VII
	Age 7	Age 11	Age 11	Age 11	Age 16	Age 16	Age 16
Top Stream	2.148 (7.986)	2.896 (19.799)	2.848 (18.385)	-	1.818 (14.209)	1.472 (10.770)	-
Middle Stream	0.530 (1.336)	1.153 (6.953)	1.120 (6.492)	-	-0.012 (-0.092)	0.086 (0.613)	-
Bottom Stream	-2.546 (-7.260)	-3.080 (-18.701)	-2.916 (-16.036)	-	-3.003 (-19.735)	-2.535 (-15.171)	-
Wave 1 reading score	-	-	0.159 (2.894)	0.112 (1.695)	-	0.624 (11.200)	0.701 (6.880)
Constant	35.978 (712.746)	35.237 (592.107)	35.230 (560.895)	35.230 (543.641)	26.426 (265.504)	26.443 (250.565)	26.437 (265.160)
Sample Size	12596	12045	10882	7214	9809	8690	1866
Adjusted R-squared	0.009	0.070	0.072	0.0004	0.108	0.117	0.024
P-value	0.000	0.000	0.000	0.090	0.000	0.000	0.000

Notes:

OLS regressions; dependent variable is class size. In age 16 regressions, the dependent variable is class size for English lessons, and streams refer to streams for that subject. Students in unstreamed classes form the omitted category for ability group.

Sample restricted at age 7 and 11 to students in state infant, junior or combined schools with 20 or more pupils, in classes of between 20 and 45 students. Sample restricted at age 16 to students in comprehensive, secondary modern, and grammar schools with 40 or more pupils, in classes of between 15 and 40 students. In columns IV and VII, sample restricted to students in unstreamed classes.

T-statistics in parentheses.

Normalised Wave 1 reading scores are used in regressions III, IV, VI and VII – hence, the coefficients denote the difference in class size associated with 1 standard deviation in reading scores.

4.1 Approaches to the problem

Blundell et al. (1997) identify three approaches to the problem of endogenous school resources. One approach controls for family fixed effects using data from siblings – as used, for example, by Altonji and Dunn (1995 and 1996). Since the NCDS contains data on only around 150 sibling pairs (all of whom are twins or triplets) this approach is not an option here.

The second approach to which they refer is that of instrumental variables, which involves finding an instrument which is correlated with class size but not with ability. This is the approach taken here.

However, this paper also makes use of the third approach identified by Blundell et al: a 'matching' approach. This involves 'washing out' the problem of unobserved ability by including in regressions as wide as possible a range of background variables. The NCDS contains an enormous range of such variables[10].

Several authors have used local-authority based variables to instrument class sizes. Feinstein and Symons (1999) use a set of local authority dummies to instrument the pupil-teacher ratio. Dolton and Vignoles (1998) and Dearden, Ferri and Meghir (1997) use variables at the local authority level, such as average pupil-teacher ratio in the LEA, or educational spending per pupil in the LEA. All these instruments suffer from the problem of being correlated with the socio-economic composition of the region of residence, as discussed before.

Akerheilm (1995) uses average class size in the student's school, and total enrolment as instruments for class size. However, this relies on the rather strong assumption that school size is not related to educational attainment.

Angrist and Lavy (1999) also use the relationship between school size and class size to identify their estimates, but are able to make use of non-linearity and non-monotonicity in this relationship, arising from a rule laid down in the Torah (Maimonides' rule) and

[10] For example, we have available data on whether the child was breast fed as an infant, on maternal smoking during pregnancy, laterality, and head circumference, all of which have been shown in the medical/child development literature to be related to some measure of 'intelligence'. There is some debate as to whether these relationships have a biological basis or whether they simply measure aspects of social class which other measures do not pick up. For our purposes, it is unimportant which of these explanations is true.

still currently in use, stipulating a maximum class size of 40 students. This gives rise to a predicted 'saw-toothed' relationship between school enrollment and class size, with discontinuities at 41, 81, 121, and so on. These predicted class sizes fit the actual data reasonably well; using this instrument, Angrist and Lavy find that smaller classes are beneficial in terms of educational attainment, having an effect comparable to (although at the lower end of) that found in the Tennessee experimental studies.

Would the NCDS permit the use of such an instrument? Figure 4 plots the relationship at age seven between class size and school size for the two types of school attended by NCDS children: infant schools, taking children from 5-7, and combined schools, taking children from age 5-11. Inspection of the graph reveals no saw-toothed pattern such as exists in the Israeli data, but it does indicate an alternative instrument: class size and school size are related for both types of school, but for any given size of school, average class sizes in infant schools are larger than in combined schools[11].

[11] For this cohort, the intake into primary schools was staggered, with most children starting school at the beginning of the term in which they would reach the age of five, and remaining in primary school until age 11. About half the children attended an 'infant' school until age seven, followed by a 'junior' school until age eleven; the other half attended a 'combined' infant and junior school for the duration of their primary schooling. In early summer (the time when most of the schools questionnaires were filled in) infant schools would have children on roll from two full year groups plus about two thirds of the youngest year group; junior schools would have children on roll from four full year groups; while combined schools would have children on roll from six full year groups plus about two thirds of the youngest group.

Figure 4: Average class size, by school size and type: NCDS Wave 1.

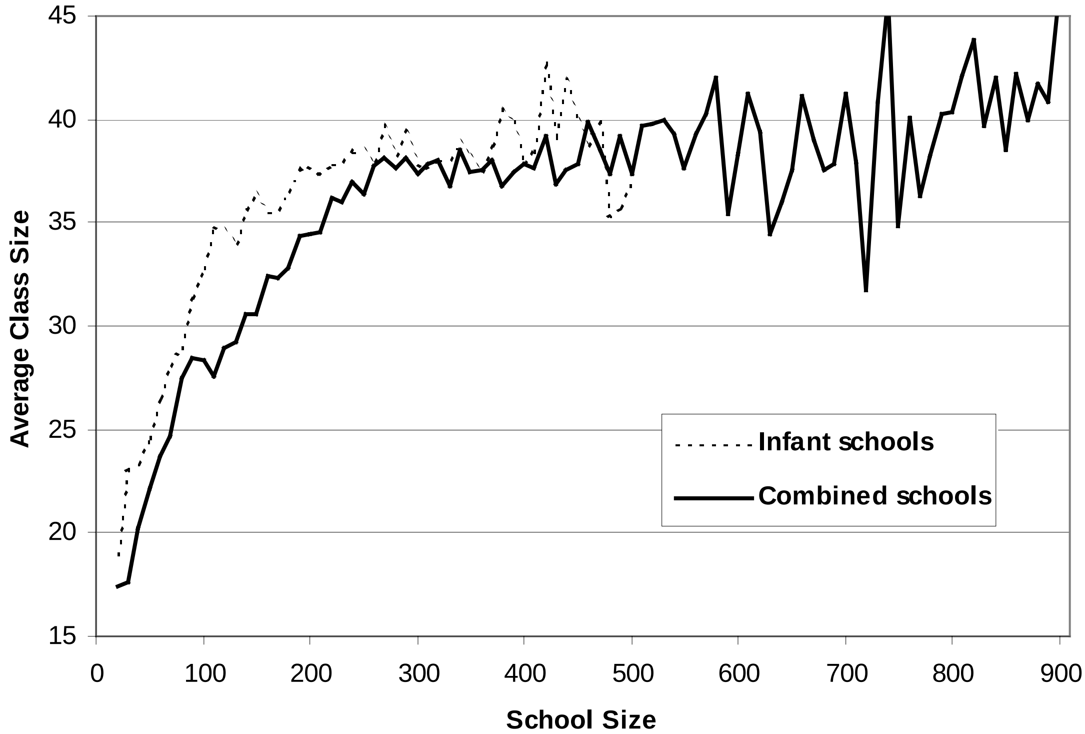


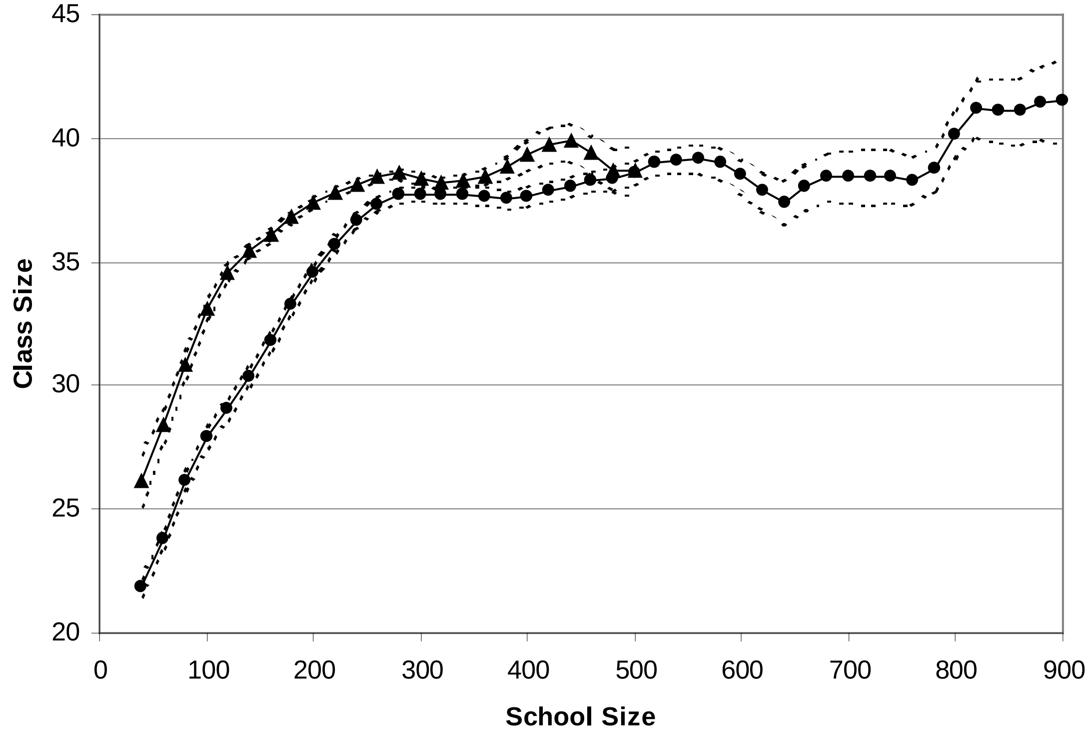
Figure 5 gives smoothed plots[12] of the relationship between school size and class size for Wave 1. Estimates for infant and combined schools are shown separately; 95% confidence intervals for both types of school are shown in broken lines. The fact that the confidence intervals lie well away from each other for most of the range confirms that there is a distinct relationship between class size and school size for the two types of school. However, Figure 6 shows that the same is not true at age 11, where the curves for the two types of school lie much closer together, and the confidence intervals (for clarity, not shown on the graph) overlap[13].

[12] These plots are kernel densities, obtained using a smoothing bandwidth of 20, excluding from the estimation process schools with under 20 pupils or over 950 pupils, and excluding those in classes of under 15 or over 50. For combined schools, estimates were obtained for schools with rolls between 40 and 900 pupils, cutting off 3% of the smallest and 0.39% of the largest schools; for infant schools, they were obtained for schools with between 40 and 500 pupils, cutting off 0.34% of the smallest and 0.64% of the largest schools.

[13] One possible reason why the relationship between school size and class size is distinct between school types at age 7 but not at age 11, is that at age 7 there is a large difference between the number of year groups in the different types of school (two and two thirds for infant schools to six and two thirds for combined schools), while at age 11, the difference is much smaller (four years for junior schools versus six and two thirds for combined schools).

Regrettably, then, this method does not furnish an exogenous instrument for class size at age 11; however, it *does* provide one for use at age 7. Although it is quite likely that both the size and type of school are related to student outcomes, the *interaction* between these two variables may be used as an exogenous instrument for class size. The only assumptions that need be made are firstly that the interaction terms are related to class size (which they are, as we shall show) and secondly that they are *not* related to student attainment. This second assumption is equivalent to assuming that school size and school type both may affect student performance, but that their effects are independent: in other words, being in a school with 500 rather than 300 students has the same effect on performance regardless of whether the school is an infant or a combined school; and the effect of being in an infant or a combined school does not vary according to whether it is a large or a small school.

The relationship between class size, school size and school type: kernel regressions for Wave 1.



The relationship between class size, school size and school type: kernel regressions for Wave 2.

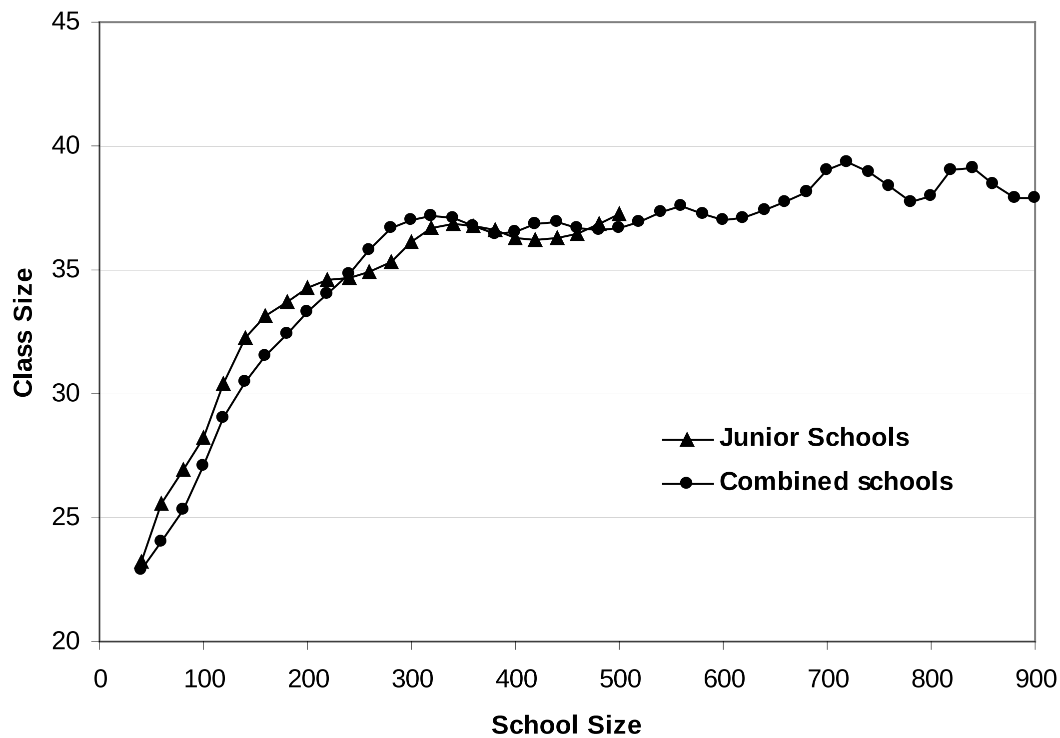


Table 4 shows that in an OLS regression with class size as the dependent variable, the interaction terms between school size and school type do have some explanatory power over and above that provided by other school-based variables; moreover, the explanatory power of the instrumenting equations is high. Column I in the table shows regression coefficients with controls only for school size, school type, whether the school has a nursery class, and the composition of the student's class. In Column II, the two interaction terms have been added: both are significant at the 1% level. In Column III, the average class size in the student's local authority has also been added in (this is used in a specification where local authority fixed effects are included); again, this coefficient is significant at the 1% level.

TABLE 4: OLS COEFFICIENTS

DEPENDENT VARIABLE: CLASS SIZE AT AGE 7 (WAVE 1)

	(I)	(II)	(III)
School Characteristics			
Infant school dummy	2.380	2.992	2.931
	(24.973)	(6.955)	(6.925)
School roll	0.058	0.065	0.062
	(43.949)	(39.152)	(37.356)
School roll^{2} (x 100)	-0.006	-0.011	-0.010
	(-28.575)	(-20.156)	(-19.429)
Infant school School roll	-	0.010	0.009
		(3.108)	(2.997)
Combined school School roll^{2} x 100	-	-0.005	-0.004
		(7.696)	(7.633)
School has nursery class	-1.029	-1.002	-0.945
	(-6.692)	(-6.594)	(-6.324)
Class Characteristics			
Top stream	0.962	0.822	0.646
	(4.060)	(3.514)	(2.804)
Middle Stream	-0.813	-0.536	-0.778
	(-2.368)	(-1.577)	(-2.324)
Bottom Stream	-4.428	-4.394	-4.516
	(-14.582)	(-14.642)	(-15.290)
All infants in one class	-1.622	-0.601	-0.521
	(-6.334)	(-2.225)	(-1.960)
Family Groupings	-0.966	-0.698	-0.712
	(-4.849)	(-3.530)	(-3.661)
Characteristics of LEA			
LEA pupil/teacher ratio	-	-	0.471
			(18.522)
Constant	24.947	23.254	10.422
	(120.678)	(85.415)	(14.031)
Observations	10399	10399	10399
R-squared	0.345	0.362	0.383
Adjusted R-squared	0.345	0.362	0.382
P-value	0.000	0.000	0.000

Note: Sample restricted to students in state infant and combined schools with 20 or more pupils, in classes of 20-45. T-statistics in parentheses. LEA average pupil-teacher ratios taken from CIPFA (1965).

5 ESTIMATION AND RESULTS

5.1 An education production function

To formalize the discussion of the previous section, we wish to estimate a reduced-form education production function of the form

$$\boldsymbol{T_i} = \boldsymbol{\alpha}.\boldsymbol{X_i} + \boldsymbol{\gamma}.\boldsymbol{S_i} + \boldsymbol{u_i} \qquad (1)$$

where $\boldsymbol{T_i}$ denotes the test score achievement of student *i;* $\boldsymbol{X_i}$ is a vector of variables including the child's personal characteristics; a set of inputs from the child's home environment; and a set of inputs (other than class size) from the child's school. $\boldsymbol{S_i}$ refers to the child's class size, and $\boldsymbol{u_i}$ is an i.i.d. error term, including a component of unobservable ability.

Some of the variables in $\boldsymbol{X_i}$ may be thought of as fixed; others may arise from the optimization of a household utility function by the child's parents. At the level of the household, for example, variables such as the total number of children, or the amount of interest parents take in their children, may be determined in this way. Some variables at the level of the school may also be determined via optimizing behaviour on the part of parents; however, given the strict operation of catchment areas at the time, it is likely that this occurs via parents choosing their area of residence rather than a particular school.

In any case, the important assumption for this model is that the set of background variables is sufficient to control for these aspects of parental choice:

$$\boldsymbol{E}(\boldsymbol{X_i}.\boldsymbol{u_i}) = 0$$

However, the previous discussion has shown that this is not the case for class size, which is related to the error term because of allocation mechanisms at both the school and the local authority levels. Hence,

$$\boldsymbol{E}(\boldsymbol{S_i}\boldsymbol{u_i}) \neq 0$$

OLS estimates are therefore biased, and we estimate instead an instrumental variables specification, using as instruments for class size the interactions between a quadratic in school size and a binary indicator of school type. In order to leave all the required variables in the main equation, the instrument set ***Z*** contains just two variables:

$$\boldsymbol{Z} = \boldsymbol{\{}[\textit{School roll} * (\textit{Infant school} = 1)]\boldsymbol{,}\ [\textit{School roll}^2 *(\textit{Infant school} = 0)]\ \boldsymbol{\}}$$

The identifying assumptions are

$$\boldsymbol{E}(\boldsymbol{Z}_i.\boldsymbol{S}_i) \neq 0 \quad ; \quad \boldsymbol{E}(\boldsymbol{Z}_i.\boldsymbol{u}_i) = 0$$

A second specification is also estimated, allowing for fixed effects at the level of the local authority: here, the error term is parameterized by

$$\boldsymbol{T}_{ij} = \boldsymbol{\alpha}.\boldsymbol{X}_{ij} + \boldsymbol{\gamma}.\boldsymbol{Z}_i + \boldsymbol{u}_{ij} \qquad (2)$$

$$\boldsymbol{u}_{ij} = \boldsymbol{\mu}_j + \boldsymbol{\varepsilon}_i$$

and the fixed effect is controlled for using a set of dummies for local authority of residence. In this specification, average class size at the level of the local authority is added to the set of exogenous instruments for class size.

A Tobit specification is also estimated for reading scores. The reading test administered to seven-year-olds, being designed to identify 'problem' readers, does not follow a bell-type distribution, but has a clustering of scores at the top, with 18% of the sample scoring the top mark of 30. The Tobit specification specification is estimated once using a straightforward Tobit procedure (corresponding to OLS in the first two specifications), and a second time using predicted class size instead of actual class size as a dependent variable (corresponding to IV in the first two specifications).

5.2 Results

Results from the first specification are shown in Table 5. Under OLS, class size appears to be unrelated to reading scores, with a tiny and insignificant negative coefficient on class size. Under IV, however, a relationship between class size and reading scores is observed which is not only significant at the 1% level, but which is also of a sizeable magnitude: -0.036 standard deviations for each reduction in class size of one pupil. Thus an eight-student reduction in class sizes corresponds to an improvement in test scores of 0.288σ: this is equal to the advantage that girls enjoy over boys in reading at this age; it is slightly larger than the difference between the performance of the top three social classes over the bottom social class; it is ten times the size of the effect of an extra year's education for the child's mother, and in terms of a peer group effect it is equivalent to the effect of moving from a class with 70% of children in the lowest social class and the rest in the middle groups, to a class with 70% of children in the *top* two social classes and the rest in the middle groups. The other coefficients are well-defined and within the bounds of reasonable expectations: children in top streams perform more than one standard deviation better than those in bottom streams; all the indicators of privileged family background are associated with better outcomes, including factors such as education, social class, fewer rather than more siblings, parental interest and so on. Some measurements from the NCDS medical questionnaire which are also known to be related to educational outcomes (height and head circumference) are also included, and they too yield significant coefficients of the expected sign. Area dummies are included for only two regions: Wales, which being very sparsely populated tends to have rather smaller classes than the rest of the country, and London, which being densely populated has rather larger classes. These variables are not significant in the OLS regression, though the dummy for London is significant in the IV regression.

TABLE 5:
DETERMINANTS OF READING SCORES: OLS AND IV ESTIMATES

		OLS		IV	
School Characteristics					
Class size		-0.0003	(-0.185)	-0.036	(-2.937)
Number on school roll		-0.0004	(-1.479)	0.0014	(2.074)
School roll squared (coefficient * 1000)		0.0004	(0.844)	-0.0015	(-1.983)
Infant school dummy		0.043	(2.198)	0.128	(3.647)
Class formation:	Top stream	0.409	(12.648)	0.439	(12.769)
	Middle stream	0.199	(3.630)	0.166	(2.997)
	Bottom stream	-0.533	(-7.894)	-0.687	(-7.812)
	Family Grouping	-0.222	(-4.781)	-0.260	(-5.454)
	All infants in one class	-0.042	(-0.810)	-0.107	(-1.868)
	School has nursery class	-0.071	(-2.224)	-0.100	(-2.959)
% of parents in social class I and II in class		0.001	(2.059)	-0.002	(2.855)
% of parents in social class V in class		-0.002	(-3.909)	-0.002	(-3.542)
% of sessions absent this year		-0.009	(-9.005)	-0.009	(-8.933)
Family Variables: Early History					
Female Child		0.280	(16.043)	0.288	(16.105)
Multiple birth indicator		-0.164	(-2.621)	-0.152	(-2.369)
Child breastfed after age 1		0.043	(2.029)	0.044	(2.063)
Age mother left education		0.030	(4.618)	0.028	(4.221)
Age father left education		0.013	(2.465)	0.010	(1.930)
Parental and Family Characteristics					
Father's social class:	I	0.231	(4.157)	0.247	(4.363)
	II	0.222	(4.597)	0.222	(4.552)
	III (non-manual)	0.280	(5.747)	0.280	(5.744)
	III (manual)	0.121	(2.790)	0.121	(2.778)
	IV	0.033	(0.705)	0.036	(0.757)
Number of children in household		-0.053	(2.977)	-0.053	(-8.295)
Child ever in care?		-0.216	(-2.845)	-0.243	(-3.121)
Family are owner occupiers		0.055	(2.889)	0.062	(3.171)
Parental aspirations					
Parents want child to stay on at school		0.231	(4.531)	0.249	(4.784)
Teacher's assessment:	Mother's interest (0-2)	0.263	(12.652)	0.263	(12.456)
Teacher's assessment:	Father's interest (0-2)	0.154	(7.086)	0.162	(7.274)
Variables from medical questionnaire					
Child's height (inches)		0.021	(5.211)	0.019	(4.672)
Child's head circumference (inches)		0.519	(2.567)	0.514	(2.640)
Head circumference squared		-0.011	(-2.325)	-0.011	(-2.372)
Area Dummies					
School in Wales		-0.011	(-0.302)	-0.096	(-2.037)
School in Inner London		-0.072	(-1.709)	-0.057	(-1.351)
Observations		11057		11057	
R-squared		0.2435		0.2202	
P-value		0.0000		0.0000	
t-statistic for endogeneity of class size				2.971	

Notes: Sample includes children in mainstream state schools with over 20 pupils, in classes of between 20 and 45. Dependent variable is normalized reading score at age 7. T-statistics in parentheses. Instruments for IV regressions are the interaction terms between school size and school type. Additional controls for father unemployed; no father figure; mother and father 'too interested' in education; child breastfed till age 1; number of schools attended; also dummies for missing values of variables. Omitted categories in groups of categorical variables are: father's social class V; combined school; unstreamed class with 1 or more classes per year. Descriptive statistics and instrumenting equations are given in Appendix 1.

TABLE 6:
READING SCORES: OLS AND IV RESULTS FROM 3 SPECIFICATIONS

	Specification I		Spec II		Specification III	
	OLS	IV	OLS	IV	Actual class size	Predicted class size
Class size coefficient	-0.0003	-0.0360	-0.0011	-0.0363	-0.0014	-0.0282
T-statistic on class size coefficient	(-0.185)	(-2.937)	(-0.529)	(-2.752)	(-0.622)	(-2.201)
T-statistic on endogeneity of class size		2.971		2.714	-	2.960
P-value (joint significance of regional dummies)			0.000	0.000		
Sample Size	11057	11057	11033	10033	10057	11057
R-squared	0.244	0.220	0.269	0.255	-	-
R-squared (instrumenting equation)	-	0.370	-	0.433	-	0.370
P-value (joint significance of all coefficients)	0.000	0.000	0.000	0.000	0.000	0.000

Notes: Dependent variable is normalized reading score at age 7. Sample restricted to students in state infant and combined schools, in classes of between 20 and 45 students.

Specification I is the same as in Table 5, using two interaction terms between school size and school type as instruments; Specification II includes a set of LEA dummies as explanatory variables, and average LEA pupil-teacher ratio as an additional instrument. Specification III is the same as Specification I except that the first estimate is run as a Tobit and the second as a Tobit with actual class size replaced by predicted class size.

T-statistics based on robust standard errors as proposed by White (1980); T-statistics on endogeneity of class size are obtained via an augmented regression procedure as suggested by Davidson and McKinnon (1993).

TABLE 7:
OLS AND IV RESULTS USING PERCENTILE AS DEPENDENT VARIABLE

	Specification I		Spec II		Specification III	
	OLS	IV	OLS	IV	Actual class size	Predicted class size
Class size coefficient	-0.0359	-1.0176	-0.0628	-1.089	-0.0014	-0.0282
T-statistic on class size coefficient	(-0.659)	(-2.963)	(-1.064)	(-2.930)	(-0.622)	(-2.201)
T-statistic on endogeneity of class size		2.926		2.825	-	2.960
P-value (joint significance of regional dummies)			0.000	0.000		
Sample Size	11057	11057	11033	10033	10057	11057
R-squared	0.249	0.226	0.270	0.248	-	-
R-squared (instrumenting equation)	-	0.370	-	0.433	-	0.370
P-value (joint significance of all coefficients)	0.000	0.000	0.000	0.000	0.000	0.000

Notes: Dependent variable is normalized reading score at age 7. Sample restricted to students in state infant and combined schools, in classes of between 20 and 45 students.

Specification I is the same as in Table 5, using two interaction terms between school size and school type as instruments; Specification II includes a set of LEA dummies as explanatory variables, and average LEA pupil-teacher ratio as an additional instrument. Specification III is the same as Specification I except that the first estimate is run as a Tobit and the second as a Tobit with actual class size replaced by predicted class size.

T-statistics based on robust standard errors as proposed by White (1980); T-statistics on endogeneity of class size are obtained via an augmented regression procedure as suggested by Davidson and McKinnon (1993).

TABLE 8:
MATHEMATICS SCORES: OLS AND IV RESULTS FROM 2 SPECIFICATIONS

	Specification I		Spec II	
	OLS	IV	OLS	IV
Class size coefficient	0.0002	0.0040	-0.0000	-0.0035
T-statistic on class size coefficient	(0.105)	(0.328)	(-0.012)	(0.253)
T-statistic on endogeneity of class size		-0.385		0.207
P-value (joint significance of regional dummies)			0.0000	0.0000
Sample Size	11032	11032	10012	10012
R-squared	0.132	0.132	0.155	0.155
R-squared (instrumenting equation)		0.370		0.434
P-value (joint significance of all coefficients)	0.0000	0.0000	0.0000	0.0000

Notes: Dependent variable is normalized mathematics score at age 7. Sample restricted to students in state infant and combined schools, in classes of between 20 and 45 students.

Specifications I and II are the same as in Table 6. Specification I uses two interaction terms between school size and school type as instruments; Specification II includes a set of LEA dummies as explanatory variables, and average LEA pupil-teacher ratio as an additional instrument.

T-statistics based on robust standard errors as proposed by White (1980); T-statistics on endogeneity of class size are obtained via an augmented regression procedure as suggested by Davidson and McKinnon (1993).

These estimates of the effect of class size (0.288 standard deviations for a reduction in class size of 8 pupils) do appear to be rather large. However, they are certainly in the same ball park as results from other research. For a reduction of on average 8 students per class, Krueger (1999) suggests effect sizes of 0.20, 0.28, 0.22 and 0.19 standard deviations in kindergarten, first grade, second grade and third grade respectively. Angrist and Lavy (1999) suggest a slightly smaller effect size of 0.18 standard deviations for fifth graders, and an effect size about half this size for fourth graders. In other words, these estimates are on the high end of what others have found, but are by no means outlandishly large.

Table 6 presents results from two further specifications; both give similar results in terms of the estimated effect of class size. The second specification (including a set of local authority dummies plus LEA average pupil teacher ratios as an extra instrument for class size) yields an almost identical coefficient on class size. The third specification is a Tobit regression intended to allow for a clustering of high values in the reading test scores; again, the simple Tobit model shows no association between class sizes and student outcomes, while the same model using predicted rather than actual class size yields a significant negative coefficient.

Turning now to the effects of class size on mathematics scores, a rather different story emerges. Results for two specifications corresponding exactly to the first two reported for reading scores are tabulated in Table 8 (no Tobit specification was necessary as the mathematics test scores follow a bell-shaped distribution). As before, under OLS no relationship is observable between class size and student attainment. However, in the case of mathematics scores, using the IV estimator makes no difference at all. No relationship between class size and attainment emerged under either specification.

Why should this be? As Robertson and Symons (1996) note, it is a good deal harder to explain attainment in mathematics than in reading: the R-squared statistics when looking at mathematics scores are typically around half the size of those found when looking at reading scores. This may possibly be because mathematics ability is somehow more ‘innate’ than reading ability, which is more easily passed on either by a favourable home environment or by an effective school. However, this conflicts with the findings of Case and Deaton (1999), who find an effect on mathematics attainment from smaller classes around five times the size of the effect on reading attainment. An alternative explanation is that the NCDS maths test administered at age 7 simply did not measure attainment in mathematics very well. This is supported by the finding that mathematics scores at ages 11 and 16 are much more closely associated with reading scores at age 7 than with mathematics scores at age 7.

5.3 Testing for heterogeneous effects

Heterogeneity is investigated along four possible axes: whether the effects of class size differ between boys and girls; between those in more and less privileged social groups; and between children in different-sized families. Additionally, we investigate whether the effect of per-pupil reductions in class size are the same for large classes as for classes which are small to begin with.

Heterogeneous effects between groups is investigated by adding to the regressions an interaction term between the predicted class size variable and the variable of interest (e.g, class size * 'girl' dummy), and testing for the significance of this extra variable. This approach assumes that the instrumenting equations and all coefficients except that on class size are identical between groups (for example, between boys and girls) and the only coefficient allowed to vary between groups is that on class size. This assumption is then relaxed by estimating separate IV regressions for the different groups (eg, for boys and girls) and testing whether the class size coefficients differ between the two regressions.

To test whether the effect of class size is different at different levels of class size, predicted class size is interacted with a dummy taking the value 1 if predicted class size is 30 or less; additionally, another specification is estimated with a quadratic term in predicted class size.

Results from the tests for heterogeneity are reported in Table 9. These show no evidence at all of any heterogeneous effects. For both reading and mathematics scores, whether just one or all the coefficients are allowed to vary, for all the groups of interest, the class size coefficients are very close together for both groups, with no significant difference between them. The conclusion from this is that reducing class sizes has a significant effect on reading scores for *all* the groups we have examined, while there is no discernible effect for *any* group in terms of mathematics scores. In regard of mathematics attainment, there is no group for which a significant effect of class size is being masked by the lack of effect for another group.

Why should this lack of heterogeneity be observed in the NCDS when it was such a feature of the Project STAR findings? One reason may be that in 1960s Britain, inequality of opportunity did exist, but not to the extent that it did in Tennessee in the 1980s. Alternatively, it may be that the level of inequality was just as high for the British sample, but not in a way which may be captured neatly by a single variable, such as ‘race’ in the STAR data.

As well as finding no evidence for heterogeneity between groups, there is no evidence that the effect of reducing class size is stronger in one part of the range of class sizes than another. Neither the regression including an interaction effect, nor the regression including a quadratic term in predicted class size produces any significant results to suggest that reducing class size has a different effect according to whether classes are small or large to begin with.

TABLE 9:
TESTING FOR HETEROGENEOUS EFFECTS

		Class size coefficient allowed to vary		All coefficients allowed to vary	
		Interaction coeff.	T-statistic	Coefficient	T-statistic
Reading	Girl	-0.0035	(-0.693)	-0.0310	(-2.026)
	(Boy)			*-0.0400*	*(-2.053)*
	Social class 1, 2 or 3 (non-manual)	-0.0006	(-0.134)	-0.0396	(-2.101)
	(Social class 3 (manual), 4 or 5)			*-0.0349*	*(-2.291)*
	Family of 3 or more children	0.0009	(1.393)	-0.0377	(-2.084)
	(Family of 1 or 2 children)			*-0.0323*	*(-2.025)*
	Predicted class size is 30 or less	0.0028	(1.452)		
	Predicted class size	-0.0776	(-1.574)		
	Predicted class size squared	0.0006	(0.864)		
Maths	Girl	-0.0015	(-0.284)	-0.0033	(-0.205)
	(Boy)			*0.0141*	*(0.742)*
	Social class 1, 2 or 3 (non-manual)	0.0052	(0.869)	0.0135	(0.538)
	(Social class 3 (manual), 4 or 5)			*0.0020*	*(0.0139)*
	Family of 3 or more children	0.0005	(0.523)	-0.0168	(-0.981)
	(Family of 1 or 2 children)			*0.0275*	*(1.500)*
	Predicted class size is 30 or less	0.0013	(0.631)		
	Predicted class size	0.0253	(0.492)		
	Predicted class size squared	-0.0003	(-0.399)		

5.4 Do the effects of smaller classes persist?

A further issue of interest from the policy perspective is whether or not the positive effects of being in a small class in infant school persist beyond the infant years. The Lasting Benefits Study (Nye et al., 1994), following up from Project STAR, found that the benefits of being in a small class in the early years persisted beyond elementary school into the first years of high school (no findings beyond this point have yet been reported). For the NCDS sample, the persistence (or otherwise) of the positive effect is estimated using OLS regressions, with test scores at age 11 and 16 as the dependent variable. Three different specifications were tested. The first is a reduced form specification, using the same set of explanatory variables as the age 7 regressions. All explanatory variables are measured at age 7 and no information after this age is included in the regression. A second specification includes a set of school-level explanatory variables measured at ages 11 and 16. A third specification includes test scores at age 7 as additional controls. All these specifications were tested firstly on the full sample, and then on the sample broken into the categories outlined in the previous section on heterogeneity.

The sample is restricted to those who were in mainstream state schools in Waves 1 and 2 (or Waves 1 and 3 for the age 16 regressions), and not in over-sized or under-sized classes. Additionally, it is restricted to children who remained in the same local education authority between the different waves, since this removes the need to control for two sets of fixed effects, as well as the possibility that ‘stayer’ and ‘mover’ families may be different from one another.

Results from the third specification which includes test scores at age 7 as additional controls are not reported. Under this specification, the coefficients on both actual class size (under OLS) nor predicted class size (under two-stage least squares) are tiny in magnitude and insignificant, leading to the conclusion that any persistent effect from being in a smaller class may be completely explained by its effect on test scores at age 7, and that there is no discernible effect additional to this.

TABLE 10:
DO THE EFFECTS OF CLASS SIZES PERSIST? EVIDENCE FROM WAVE 2

		Reading		Maths	
		Reduced form	With Wave 2 controls	Reduced form	With Wave 2 controls
All	OLS	0.0001	-0.0004	0.0000	-0.0009
		(0.0431)	(-0.2119)	(-0.0656)	(-0.4839)
	2SLS	-0.0215	-0.0101	-0.0196	-0.0142
		(-1.5256)	(-0.7387)	(-1.3179)	(-1.0366)
Girls	OLS	-0.0006	-0.0013	-0.0030	-0.0035
		(-0.2187)	(-0.4744)	(-0.9730)	(-1.1824)
	2SLS	**-0.0345**	**-0.0355**	-0.0330	**-0.0429**
		(-1.746)	**(-1.8227)**	(-1.5986)	**(-2.2009)**
Boys	OLS	0.0004	0.0002	0.0023	0.0012
		(0.1104)	(0.0744)	(0.7111)	(0.4063)
	2SLS	-0.0106	0.0155	-0.0076	0.0142
		(-0.5274)	(0.7801)	(-0.3568)	(0.7229)
Big family (3+ children)	OLS	0.0023	0.0009	0.0022	0.0000
		(0.7857)	(0.3055)	(0.7689)	(-0.0013)
	2SLS	-0.0263	-0.0161	**-0.0330**	**-0.0329**
		(-1.3767)	(-0.8642)	**(-1.696)**	**(-1.8161)**
Small family (1 or 2 children)	OLS	-0.0031	-0.0025	-0.0039	-0.0029
		(-0.9239)	(-0.7602)	(-1.0906)	(-.8642)
	2SLS	-0.0182	-0.0084	-0.0043	0.0039
		(-0.8494)	(-0.4039)	(-0.1898)	(0.1830)
Social Class I, II and III (n/m)	OLS	-0.0009	-0.0021	-0.0011	-0.0028
		(-0.3864)	(-0.9001)	(-0.4784)	(-1.2468)
	2SLS	-0.0196	-0.0096	-0.0202	-0.0188
		(-1.2008)	(-0.6226)	(-1.1973)	(-1.2295)
Social Class III(m), IV and V	OLS	0.0054	0.0068	0.0030	0.0044
		(1.1092)	(1.4100)	(0.5980)	(0.9089)
	2SLS	-0.0247	-0.0085	-0.0174	0.0034
		(-0.8944)	(-0.2912)	(-0.5591)	(0.1140)

Notes: Figures reported are coefficients on class size (or predicted class size) from OLS and 2SLS regressions with normalized reading and mathematics score at age 11 as dependent variable. Robust T-statistics are in parentheses. Sample restricted to students in mainstream state schools, in classes of between 20 and 45 students at Waves 1 and 2; and those who have not moved between LEAs between the two waves. For two-stage least squares, the predicted class size variable is as calculated in the sample used in Tables 5 and 6.

Results at age 11 from the other two specifications are shown in Table 10. The table contains four columns of results: the first two from regressions with reading scores at age 11 as the dependent variable, and the other two from regressions with mathematics scores at age 11. For each subsample, the coefficient on the class size variable is shown for OLS and two-stage least squares regressions. A general feature of these results is that under OLS the coefficient on class size is always close to zero and never significant, while for several groups under two-stage least squares, the coefficient is negative, much larger in magnitude, and in some of the regressions, significant at the 10% or even the 5% level. The groups for whom class size at age 7 has an effect which persists until age 11 are girls (for whom the coefficients on reading scores are negative and significant at the 10% level, and for whom the coefficient on mathematics score is significant at the 5% level in the specification including Wave 2 controls); and children from larger families, for whom the coefficients on mathematics scores are negative and significant at the 10% level. At the time, these might be thought of as having been 'disadvantaged' groups; interestingly, though, the effect of small class sizes does not appear to persist for members of the other 'disadvantaged' group, namely children from the lower socio-economic classes.

The size of the coefficients for the groups where persistence is evident is between -0.03 and -0.04, similar to the size of coefficients in the regressions for age 7, but if anything slightly larger. Hence, for the girls and for children with two or more siblings in this sample, a reduction in class size by eight students at age 7 would be associated with an increased performance in test scores at 11 of between 0.24 and 0.32 standard errors.

The analysis just discussed was repeated to examine the relationship between class size at age 7 and test scores at age 16. In some ways, a similar pattern is visible: namely, that under OLS the coefficient on class size tends to be small and insignificant, and is usually positive, whereas under two-stage least squares the coefficient is more often negative and larger. However, even under 2SLS, the effect is never significant. Perhaps this is due to the smaller sample sizes in these regressions; perhaps it is due to the greater time lag; or perhaps the effects of small infant class sizes really do not persist until age 16.

6 CONCLUSIONS

This paper has estimated the effects of class size on student test scores, using as an exogenous instrument for class size the interaction between school size and school type, thus freeing the estimates of bias arising from the redistributive allocation of educational resources faced by the children in this sample.

OLS estimates of the relationship between class size and student attainment have tended to yield insignificant and/or ‘wrongly’-signed estimates of the effect of class size. However, the instrumental variables estimates obtained in this paper indicate that when endogeneity of class size is accounted for, class size in the early years is strongly related to children’s test scores in reading. Estimated effect sizes are around 0.288 standard deviations for a reduction in class size of 8 pupils, and these are comparable with the results found by Finn and Achilles (1990), Angrist and Lavy (1999), and Krueger (1999).

Rather surprisingly, class size was *not* found to have a significant effect on mathematics scores. This runs counter to the findings of Case and Deaton (1999), who find an effect of educational resources on mathematics attainment around five times bigger than the effect on literacy attainment. Possible explanations are firstly that it is simply more difficult to explain attainment in mathematics than in reading, and that mathematics ability is somehow more ‘innate’ than reading ability; alternatively, it is possible that the mathematics test administered to the NCDS cohort at age 7 was simply not a particularly good measure of attainment in mathematics.

Another surprising finding is that there is no evidence of heterogeneous effects of class size between different groups of children: the effect of class size on reading scores appears to be slightly larger for girls than for boys, but this difference is not significant, and neither is there any significant difference between effect sizes for children from more and less advantaged groups. This is particularly interesting because Project STAR found a great deal of heterogeneity between groups, with Black children and those from inner cities benefiting more from smaller classes than other children. The apparent lack of heterogeneity in the NCDS sample may be because children in 1960s Britain faced a lower level of inequality than the Project STAR children from Tennessee in the 1980s; it may also be because social advantage and disadvantage cannot be captured neatly by a simple combination of variables for the NCDS children, whereas it is captured well by the ‘race’ variable for the Project STAR children.

Finally, the beneficial effects of smaller classes were found to persist through to age 11 for certain groups of children: girls, and children from larger families. It is difficult to extrapolate from these findings to make inferences about the effects of current practices. In the 1960s the groups for whom small classes had a persistent effect could be thought of as 'disadvantaged' groups; in particular, there was a good deal of educational debate about the attainment of girls. At the time of writing, girls are no longer seen as being at a disadvantage in primary school, and because of this it is not clear whether the effect of smaller classes would still be persistent for girls but not for boys.

In terms of policy, this paper makes one clear prediction: if money is spent to reduce infant class sizes, then children's attainment in reading will improve. The Government has already gone some way towards its stated aim of reducing class sizes to a maximum of 30; however, the beneficial effect of reducing class sizes is not confined to cutting very large classes down to 30, but would continue if classes were cut to below 30.

Although this paper makes clear predictions that cutting infant class sizes further, to below 30, would improve students' attainment, it is not my intention to say whether this would be a good use of public money or not: there may be other, more important, uses to which the available money could be put. This is a point made repeatedly by academics opposed to increasing educational investment, including Slavin (1980), Hanushek (1995), Prais (1996) and Pelzman (1997). Pelzman uses some 'back-of-the-envelope analysis' to argue that even if Card and Krueger's best estimates of the effect of school resources on earnings were correct, it would be a bad investment to increase educational expenditures, and the government would do better to reduce expenditures and hand out government bonds to students instead. A more rigorous cost-benefit analysis might also focus on the social benefits of improved schooling, as well as the private rate of return to schooling, and may come to a different conclusion altogether.

Of course, such a cost-benefit analysis lies well outside the scope of this paper. The purpose of this paper was to estimate the effects of class size on student attainment, and to free the estimates as much as possible from the confounding effects of compensatory resource allocation mechanisms. Having done this, the findings of this paper may be added to the growing pool of evidence that once the allocation of educational resources is properly controlled for, small classes really do have benefits in terms of educational outcomes.

7 APPENDIX

TABLE 11: DESCRIPTIVE STATISTICS

		Mean	S.D.	Min	Max
Dependent Variables					
Normalised reading score at age 7		0	1	-3.453	0.933
Normalised maths score at age 7		0	1	-2.114	1.966
School Characteristics					
Class size		36.075	5.433	20	45
Number on school roll		249.783	112.372	20	948
Infant school dummy		0.578	0.494	0	1
Class formation:	Top stream	0.038	0.191	0	1
	Middle stream	0.017	0.129	0	1
	Bottom stream	0.021	0.143	0	1
	Family Grouping	0.048	0.214	0	1
	All infants in one class	0.036	0.187	0	1
	School has nursery class	0.092	0.289	0	1
% of parents in social class I and II in class		23.631	18.034	0	100
% of parents in social class V in class		21.635	16.601	0	100
% of sessions absent this year		8.635	9.853	0	100
Family Variables: Early History					
Female Child		0.489	0.500	0	1
Multiple birth indicator		0.024	0.152	0	1
Child breastfed after age 1		0.427	0.495	0	1
Age mother left education		14.864	1.264	12	23
Age father left education		14.915	1.755	12	23
Parental and Family Characteristics					
Father's social class:	I	0.045	0.208	0	1
	II	0.127	0.333	0	1
	III (non-manual)	0.098	0.297	0	1
	III (manual)	0.436	0.496	0	1
	IV	0.165	0.371	0	1
Number of children in household		3.107	1.644	1	14
Child ever in care?		0.019	0.137	0	1
Family are owner occupiers		0.422	0.494	0	1
Parental aspirations					
Parents want child to stay on at school		0.787	0.409	0	1
Teacher's assessment:	Mother's interest (0-2)	1.106	0.765	0	2
Teacher's assessment:	Father's interest (0-2)	0.699	0.827	0	2
Variables from medical questionnaire					
Child's height (inches)		48.201	2.281	37	61
Child's head circumference (inches)		20.874	0.686	16	30
Area Dummies					
School in Wales		0.058	0.234	0	1
School in Inner London		0.050	0.218	0	1
Class Size Instruments					
Infant school * school roll		131.465	130.247	0	574
Combined school * school roll squared		115.496	161.512	0	948
Pupil teacher ratio in LEA (10036 observations)		28.608	1.717	17.6	37.4

Notes: Sample includes 11057 children in mainstream state schools with over 20 pupils, in classes of between 20 and 45. Statistics are for the sample used for regressions in Table 5; statistics for other samples may vary slightly. To preserve sample sizes, dummy variables were created to indicate missing values in the data. The means given here are for non-missing values.

TABLE 12:
DO THE EFFECTS OF CLASS SIZE PERSIST? EVIDENCE FROM WAVE 3.

		Reading		Maths	
		Reduced form	With Wave 3 controls	Reduced form	With Wave 3 controls
All	OLS	0.0031	0.0018	-0.0011	-0.0009
		(1.1119)	(0.7576)	(-0.4222)	(-0.4575)
	2SLS	-0.0006	-0.0102	-0.0002	-0.0082
		(-0.0429)	(-0.7774)	(-0.0164)	(-0.5788)
Girls	OLS	0.0000	0.0003	-0.0058	-0.0022
		(-0.0106)	(0.0813)	(-1.5351)	(-0.7303)
	2SLS	-0.0176	-0.0220	0.0035	0.0094
		(-0.7763)	(-1.1139)	(0.1439)	(0.4531)
Boys	OLS	0.0059	0.0034	0.0040	0.0006
		(1.4655)	(0.9616)	(1.0191)	(0.1786)
	2SLS	0.0112	-0.0030	-0.0012	-0.0239
		(0.5231)	(-0.1629)	(-0.0555)	(-1.2458)
Big family (3+ children)	OLS	**0.0060**	0.0028	-0.0011	-0.0011
		(1.6799)	(0.8282)	(-0.3562)	(-0.4151)
	2SLS	-0.0023	-0.0053	0.0005	-0.0136
		(-0.1282)	(-0.3121)	(0.0252)	(-0.727)
Small family (1 or 2 children)	OLS	-0.0015	0.0006	-0.0012	-0.0013
		(-0.3811)	(0.1813)	(-0.2795)	(-0.4014)
	2SLS	0.0125	-0.0073	0.0096	0.009
		(0.4730)	(-0.3449)	(0.3380)	(0.4019)
Social Class I, II and III (n/m)	OLS	0.0030	0.0021	-0.0022	-0.0014
		(0.9483)	(0.7422)	(-0.7585)	(-0.6098)
	2SLS	-0.0042	-0.0074	-0.0010	-0.0024
		(-0.2370)	(-0.4815)	(-0.0602)	(-0.1565)
Social Class III(m), IV and V	OLS	0.0015	-0.0007	0.0041	0.0002
		(0.3175)	(-0.1783)	(0.6102)	(0.0272)
	2SLS	0.0069	-0.0210	0.0016	-0.0337
		(0.2528)	(-0.8912)	(0.0413)	(-1.0494)

Notes: Figures reported are coefficients on class size (or predicted class size) from OLS and 2SLS regressions with normalized reading and mathematics score at age 11 as dependent variable. Robust T-statistics are in parentheses. Sample restricted to students in mainstream state schools, in classes of between 20 and 45 students at Waves 1 and 2; and those who have not moved between LEAs between the two waves. For two-stage least squares, the predicted class size variable is as calculated in the sample used in Tables 5 and 6.

8 Bibliography

Achilles, C.M. (1996)
"Summary of Recent Class-Size Research With an Emphasis on Tennessee's Project STAR and its Derivative Research Studies"
Research Summary, Center of Excellence for Research and Policy on Basic Skills, Tennessee State University, Nashville.

Akerheilm, K. (1995)
"Does Class Size Matter?"
Economics of Education Review volume XIV, pp. 229-241.

Altonji, J. G. and Dunn, T. A. (1995)
"The effects of School and Family Characteristics on the Return to Education"
NBER Working Paper no. 5072.

Altonji, J. and Dunn, T. A. (1996)
"Using Siblings to Estimate the Effect of School Quality on Wages"
Review of Economics and Statistics, vol. LXXVIII, no 4, pp. 665-671.

Amemiya, T. (1981)
"Qualitative Response Models: A Survey"
Journal of Economic Literature, vol. 19, pp. 1483-1536.

Angrist, J. and Lavy, V. (1999)
"Using Maimonides' Rule to Estimate the Effect of Class Size on Scholastic Achievement"
Quarterly Journal of Economics, vol. 114, no 2, pp. 533-575.

Bennett, N. (1998)
"Annotation: Class Size and the Quality of Educational Outcomes"
Journal of Child Pschology and Psychiatry, vol. 39, no 6, pp. 797 – 804.

Betts, J. (1995)
"Does School Quality Matter? Evidence from the National Longitudinal Survey of Youth"
Review of Economics and Statistics, vol. LXXVII, pp. 231 – 250.

Betts, J. (1996)
"Do School Resources Matter only for Older Workers?"
Review of Economics and Statistics, vol. LXXVIII, no. 4, pp. 638 – 652.

Blatchford, P. and Martin, C. (1998)
"The Effects of Class Size on Classroom Processes: 'It's a bit like a treadmill – working hard and getting nowhere fast' "
British Journal of Educational Studies, Vol. 46, no. 2, pp. 118 – 137.

Blatchford, P. and Mortimore, P. (1994)
"The Issue of Class Size for Young Children in Schools: what can we learn from research?"
Oxford Review of Education, Vol. 20, no. 4, pp. 411 – 428.

Blundell, R., Dearden, L., Goodman, A. and Reed, H. (1997)
"Higher Education, Employment and Earnings in Britain"
Institute for Fiscal Studies, 1997.

Brimblecombe, N. (1999)
"Inspecting the Inspection System"
In Dorling and Simpson (eds) *Statistics in Society* Arnold, London.

Card, D. and Krueger. A. (1992)
"Does School Quality Matter? Returns to Education and the Characteristics of Public Schools in the United States"
Journal of Political Economy, vol 100, pp. 1-40.

Card, D. and Krueger, A. (1992b)
"School Quality and Black-White Relative Earnings: A Direct Assessment"
Quarterly Journal of Economics, vol. 107, pp 151-200.

Card, D. and Krueger. A. (1996)
"School Resources and Student Outcomes: An Overview of the Literature and New Evidence from North and South Carolina"
Journal of Economic Perspectives, vol. 10, no. 4, pp. 31-50.

Card, D. and Krueger, A. (1998)
"School Resources and Student Outcomes"
Annals of the American Academy of Political and Social Science, vol. 559, pp. 39-53.

Case, A. and Deaton, A. (1999)
"School Inputs and Educational Outcomes in South Africa"
Quarterly Journal of Economics, vol. 114, no. 3, pp 1047 – 1084.

CACE (Central Advisory Council for Education: England) (1967)
"Children and their Primary Schools"
HMSO, London 1967.

CIPFA (Chartered Institute of Public Finance and Accountancy) 1965
"Education Statistics 1964-65: Actuals"
London: Statistical Information Service 1965.

CIPFA (Chartered Institute of Public Finance and Accountancy) 1970
"Education Statistics 1969-70: Actuals"
London: Statistical Information Service 1970.

Davidson, R. and McKinnon, J.G. (1993)
"Estimation and Inference in Econometrics"
Oxford University Press, New York 1993.

Davie, R., Butler, N. and Goldstein, H. (1972)
"From Birth to Seven"
Longman, London, 1972.

Day, C., Tolley, H., Hadfield, M., Parkin, E. and Watling, R. (1996)
"*Class size research and the quality of education*"
Haywards Heath, UK: NAHT.

Dearden, L., Ferri, J. and Meghir, C. (1997)
"The effect of school quality on educational attainment and wages"
Working Paper no. w98/3, Institute for Fiscal Studies, London.

DfEE (Department for Education and Employment) (1997)
"Excellence in Schools"
White paper on Education, October 1997.

DfEE (Department for Education and Employment) (2000)
"Class Sizes in Maintained Schools in England: January 2000 (provisional)"
DfEE Statistical First Release SFR 15/2000.

Dolton, P. and Vignoles, A. (1998)
"The Impact of School Quality on Labor Market Success in the United Kingdom"
TSER Schooling, Training and Transition Working Paper no. 8-98.

Feinstein, L .and Symons, J. (1999)
"Attainment in Secondary School"
Oxford Economic Papers 51, pp. 300 – 321.

Finn, J.D. and Achilles, C. (1990)
"Answers and questions about class size: a statewide experiment"
American Educational Research Journal 1990, vol. 27 no 3, pp. 557-577.

Glass, G. and Smith, M.L. (1979)
"Meta-analysis of research on class size and achievement"
Educational Evaluation and Policy Analysis vol. 1 pp 2-16.

Grogger, J. (1996)
"School Expenditures and Post-Schooling Earnings: Evidence from High School and Beyond"
Review of Economics and Statistics, vol. LXXVIII, no 4, pp. 628-637.

Hanushek, E. (1996)
"Assessing the Effects of School Resources on Student Performance: An Update"
Rochester Center for Economic Research, Working Paper no. 424.

Hanushek, E., Rivkin, S, and Taylor, L. (1996)
"Aggregation and the Estimated Effects of School Resources"
Review of Economics and Statistics, vol. LXXVIII, no 4, pp. 611-627.

Harries, Gillian (ed) (1994)
"Independent Schools Yearbook 1994-1995"
A&C Black, London.

Jackson, B. (1964)
"Streaming: an Educational System in Miniature"
London 1964

Krueger, Alan (1999)
"Experimental Estimates of Education Production Functions"
Quarterly Journal of Economics vol. 114, no 2, pp. 497 – 532.

Loeb, S. and Bound, J. (1996)
"The Effect of Measured School Inputs on Academic Achievement: Evidence from the 1920s, 1930s and 1940s Birth Cohorts"
Review of Economics and Statistics, vol. LXXVIII, no 4, pp. 653 – 663.

Morris, J. (1959)
"Reading in the Primary School"
London, Newnes 1950.

Mueller, D.J., Chase, D.I., and Walden, J.D. (1988)
"Effects of reduced class sizes in primary classes"
Educational Leadership, vol. 45, pp. 48 – 50.

Nye, Barbara A., Zaharias, J., Fulton, B., Achilles, C.M., Cain, V. and Tollett, D. (1994)
"The Lasting Benefits Study: A continuing analysis of the effect of small class size in Kindergarten through Third Grade on Student Achievement Test Scores in Subsequent Grade Levels: Seventh Grade Technical Report"

Center of Excellence for Research in Basic Skills, Tennessee State University, Nashville, Tennessee, 1994.

OECD (Organisation for Economic Co-operation and Development (1996)
"Education at a Glance"
Paris: OECD 1996.

Office for Standards in Education (OFSTED) (1995)
"Class Size and the Quality of Education"
Ofsted Publications 1995.

Pelzman, S. (1997)
"Class Size and Earnings - Correspondence"
Journal of Economic Perspectives vol. 11, no. 4, pp. 225-226.

Prais, S. J. (1996)
"Class size and learning: the Tennessee experiment - what follows?"
Oxford Review of Education, vol. 22, no 4, pp.399-414 .

Robertson, D. and Symons, J. (1996)
"Do Peer Groups Matter? Peer Group versus Schooling effects on Academic Attainment"
Centre for Economic Performance discussion paper no. 311, 1996.

Shapson, S. M., Wright, E. N., Eason, G. and Fitzgerald, J. (1980)
"An experimental study of the effects of class size"
American Educational Research Journal, vol. 17, pp. 141 – 152.

Simon, B. (1991)
"Education and the Social Order"
Lawrence and Wishart, London 1991.

Slavin, R. (1989)
"Class size and student achievement: small effects of small classes"
Educational Psychologist 1989, vol. 24 no 1, pp. 99 – 110.

Slavin, R. (1990)
"Class size and student achievement: is small better?"
Contemporary Education 1990, vol. 62 no 1, pp. 6 – 12.

White, H. (1980)
"A Heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity"
Econometrica vol. 48, pp. 817-838.

Word, E., Johnston, J., Bain, H., Fulton, B., Zaharias, J., Achilles, C., Lintz, M., Folger, J. and Breda, C. (1990a)
"The State of Tennessee's Student/Teacher Achievement Ratio (STAR) Project: Technical Report 1985 – 1990"
Center of Excellence for Research in Basic Skills, Tennessee State University, Nashville, Tennessee.

Word, E., Achilles, C., Bain, H., Folger, J., Johnston, J. and Lintz, N. (1990b)
"Project STAR Final Executive Summary: Kindergarten through 3rd grade results (1985 - 89)
Contemporary Education 1990, vol. 62 no 1, pp. 13 – 16.