British Educational Research journal (2002, 8,2). To appear.

A study of class size effects in English school reception year classes

by

Peter Blatchford
Harvey Goldstein
Clare Martin
and
William Browne

Institute of Education
University of London

Summary

This paper reviews the existing evidence on the relationship between class size and achievement for children in their first years of schooling. It then describes a large-scale longitudinal study of such children within English Local Education Authorities and presents results for achievement progress in literacy and mathematics during the reception year. Using a series of multilevel models, it is shown that there is a relationship with size of class, after various confounding factors have been allowed for, and that there are interactions between class size and initial achievement and between class size and entitlement to free school meals. It is argued that these results, especially the differential effects for different groups of children, could have important implications for educational policy.

Keywords

Class size research, educational policy, longitudinal study, multilevel model

Correspondence: Professor P. Blatchford, Institute of Education, 20 Bedford Way, London, WC1H 0AL, UK. *Email*: p.blatchford@ioe.ac.uk

Introduction

There has been a good deal of controversy over the effects of class size differences on pupils' educational attainments. In the U.K. debate has been about the negative effects of large classes, while in the U.S.A. debate has centred on the efficacy and cost effectiveness of class size reductions. There are profound policy implications, not the least because class size reduction initiatives – such as those now introduced into a number of States in the USA (Finn and Achilles, 1999), as well as the recent U.K. Government's pledge on maximum class sizes of 30 in all infant classes - are extremely costly. There are also policy implications concerning the exact size of class in relation to academic achievement. There is a common view, based on the U.S. research, that class size reductions below 20 are necessary before effects on achievement become noticeable (Blatchford and Mortimore, 1994), though this has not been tested in other countries such as the U.K. There is some consensus that class size effects are most marked in the case of the youngest children in school, in the first years after school entry (Blatchford and Mortimore, 1994), and also in the case of minority or poor children (Finn and Achilles, 1999, Molnar et al, 1999), though once again information on children in U.K. schools is not available.

The concept of class size is not straightforward, as we take up in more detail elsewhere (Blatchford, Goldstein and Mortimore, 1998). A pupil will experience class sizes varying from day to day as well as within days. The size of class recorded on a register will not be that experienced all the time by all pupils. For the purpose of the present study, however, we have used the registered class size for the three terms of the reception year. Another complication is that in many schools children will be recruited into the reception year at the beginning of each term, depending on their age, so that in general the experienced class size will increase throughout the year. In our analyses, which we report below, we use the average class size experienced by each pupil. In addition we adjust for the term of entry to school.

Much existing research on class size effects has been carried out and has been summarized, for example by Slavin (1989) and Blatchford and Mortimore (1994). More

recently a meta analysis has been carried out (Yang et al, 2000) which identifies nine existing studies of Primary school children which satisfy basic quality criteria such as collecting accurate data on class sizes, having assessment information at the start and end of the period being examined (or having adequate randomization) and having a large enough sample size. The principal conclusions (see also Goldstein and Blatchford, 1998) may be summarized as follows:

- A reduction in achievement of about 0.2 standardised units for an increase in class size of 10 pupils above a class size of 15. Some evidence for a non-linear relationship.
- Smaller class size effects with increasing age/grade.
- Evidence that class size effect is larger for disadvantaged pupils

The largest, and arguably most important, study to date is the STAR project carried out in the state of Tennessee in the late 1980s (Word et al, 1990). This was a randomized controlled trial (RCT) whereby 65 schools with at least 3 entry forms were selected and children randomly assigned to small (about 15), regular (about 25), and regular with teacher-aide classes. The study children were followed for up to 6 years. By contrast, other studies, while satisfying quality design criteria, have involved much smaller numbers of pupils and classes. Nevertheless, Yang et al (2000) showed that there was good agreement on the magnitude of class size effects among these studies.

A main point to make is that we lack in the UK research on the effects of class size differences. Such research as has been done has not employed research designs strong enough to arrive at unambiguous conclusions. It is commonly assumed that experimental research, such as the STAR project, can provide the best evidence on the causal effect of class size differences. However, a number of authors have noted weaknesses in the STAR study (see Goldstein and Blatchford, 1998 for further details and for a general discussion of the relative merits of randomized controlled trails and naturalistic studies in educational research). The principal ones are as follows:

1. ‘Contamination’ of effects among classes of different sizes within same schools. Cluster randomization would have been better
2. Zero blind because all the participants including children knew which treatment they were in
3. Lack of entry assessment to improve precision and inference details and check on randomization
4. Applicable strictly only to large schools – uncommon in Britain
5. The classes tend to be smaller than in the U.K. and are therefore not very informative about the more common class sizes experienced in British primary schools.

Despite these problems, which limit the usefulness of the STAR study, its general conclusions are in line with others and as we shall see also in line with those from the study here reported.

The Institute of Education Class Size Project was set up to answer, for the first time in the UK, questions about the effect of class size differences on pupils’ academic progress. The study had a number of features that were designed to be an improvement on previous research. In line with the limitations just cited concerning experimental designs, in the present study an ‘observational’ approach was adopted rather than an interventionist one involving random assignment, and a longitudinal design with baseline assessment was used in order to adjust for possible purposive or non-random selection of children into classes on the basis of their pre-existing achievement. The study followed a large sample of children from school entry through the infant stage, i.e., children aged 4 –7 years. It used a multi-method approach drawing on a range of sources of data, and used multi-level statistical procedures to model effects of class size differences while controlling for sources of variation that might affect the relationship with academic achievement. Models are fitted that allow for the hierarchical structure of educational data, i.e., that allow for between-pupil, between-classroom and between-school variation. In using an ‘observational’ design we were able to capture the nature of the relationship between class size and achievement across the full range of observed class sizes (not just a restricted range as in the case of the STAR project), and this seems important for policy

recommendations, for example, in terms of whether certain class sizes or bands of class sizes have stronger effects than others. We also employ a more sophisticated approach to modelling the relationship between class size and achievement, than that conducted in previous research, where a simple linear relationship is often assumed, though not tested.

The attribution of causality on the basis of an observational study is clearly important, The key issue is whether the relationships observed between class size and achievement can be due to further 'confounding' factors which have not been accounted for, We have already alluded to the possibility that children are allocated to classes in such a way that the initially low achieving children enter smaller classes. If this were the case then this could induce a subsequent association between class size and achievement *solely due to the initial allocation process*, The ways in which this could occur may be subtle, involving judgements by teachers, information from parents etc. In the present study we have attempted to adjust for such effects by taking baseline measurements on entry to reception class and fitting linear and nonlinear model adjustments for these, It is possible, of course, that such measures are insufficient to account for all 'assignment' factors; we have therefore made further adjustment using free school meal eligibility which acts as a proxy for poverty. Nevertheless, there remains the possibility that some or all of the relationships that we find could be accounted for by further factors and the reader is asked to bear this in mind when interpreting the results. We shall return to this issue in the discussion.

In this paper we report results concerning the first year in English infant schools – the reception year – which allows comparisons with previous research, in which, as we have seen, effects soon after school entry have been most evident. As previous research also suggests that smaller classes are more beneficial for some groups of children, e.g., poor or minority children, we also assess whether class size effects, if they exist, are modified in any way by the characteristics of the children entering schools, in particular in terms of gender, family income and attainment on entry to school. We also provide an initial assessment of the extent to which the relationship between class size and achievement is

affected by the composition of the class, in terms of the deviation of children's scores from the average.

Method

Samples

Starting in 1996, children starting school in the reception year were recruited to the study. The study followed two large cohorts of children over the first three years of school (i.e., the whole of Key Stage 1 (KS1): Reception year, Year 1 and Year 2; 4-7 years). The first cohort started school in 1996, and the second 1997. Some 220 schools, with 368 classes and 9330 children in 8 Local Education Authorities (LEAs) were involved in Cohort 1; Cohort 2 involved a further 5 LEAs. The present analyses use only cohort 1. The research design involved random selection of schools within the participating LEAs. All children entering reception in a selected school during the year were included in the study.

Data collection

School Entry Assessment

Information was collected when pupils entered school by means of a baseline entry assessment conducted by the teacher. The procedure was the Avon Reception Entry Assessment (1996), which covers literacy and mathematics and comprises information from teacher ratings, based on classroom observations, and tasks completed by children. A measure of literacy knowledge was derived by adding for each child scores on 15 items in language, 18 in reading, 17 in writing and a test of letter identification (how many of 26 letters were recognised in terms of either name or sound), and a measure of mathematics was based on total correct out of 19 items. Training was provided for class teachers in its use.

End of Reception Year Assessment

At the end of the reception year the Literacy Baseline component of the Reading Progress Test (2000), was administered by teachers. It reflects well the literacy curriculum experienced by children in English schools, is consistent with the school entry assessment, and covers phonological awareness, concepts about print, letter knowledge, word reading, spelling/writing and a specially developed teacher administered test covering understanding of number. In the case of mathematics a teacher-administered test was devised and piloted. The final version was again designed to cover the curriculum experienced by children and covered counting, repeating patterns, comparison and matching, addition using pictures, subtraction using pictures, addition and subtraction using words, addition and subtraction using symbols, and shape recognition.

Termly Questionnaires

A termly questionnaire on class sizes and classroom activities during a half-day period was also completed each term by the class teacher. This collected detailed data on registered and 'experienced' class size (i.e., the class size on the register and the class size as experienced by children at the given survey point), and number of adults present; data on proportions of time devoted to teaching and non-teaching activities; information on within class grouping practices (size and number of groups, group composition in terms of ability, friendship and age mix, the role of adults, and curriculum, task and activity type); time in different curriculum areas; and the frequency and amount of reading activities.

Pupil Background Information

Pupil background details including age, sex, free school meal entitlement, English language fluency, previous nursery education, attendance and special educational needs were also collected.

Other information collected included pupil behaviour ratings (completed by teachers on individual pupils), structured observations of classroom interactions and information on teachers, but these are not reported here.

Results

We present results for mathematics and literacy. The variables used in the present analyses are given in Table 1.

(Table 1 here)

A number of preliminary explorations of the data suggested that the modelling of class size effects should be restricted to class sizes from 10 to 35. We have fitted for Maths, however, only class sizes from 15-33. A few very small and very large classes appeared to be anomalous and have been excluded. Class size itself is measured about an origin of 30, which partly reflects the apparent importance attached to this 'threshold' by educators and policy makers. Almost all previous studies have limited themselves to linear relationships between achievement and class size, but this may only be true over a restricted range. Fitting simple polynomial relationships may impose too-rigid constraints on the shape of the relationship, especially at the extremes. We have therefore explored various regression spline approaches whereby an underlying polynomial relationship is modified by the addition of smoothly joining local polynomials at selected 'knots'. We have also studied more general fractional polynomials but these appear to add little and have not been used. The cubic regression spline which we have used (See for example Hastie and Tibshirani, 1990) is defined as follows:

$$y_i = \sum_{h=0}^{3} \beta_h x_i^h + \alpha_1 (x_i - k_1)_+^3 + \alpha_1 (k_2 - x_i)_+^3$$

$$k_1 > k_2 \qquad (1)$$

$$\text{where } (z)_+ = \begin{cases} 0 & \text{if } z < 0 \\ z & \text{if } z \geq 0 \end{cases}$$

The knots k_1, k_2, are chosen to provide the best smoothed relationship over the range of class sizes fitted. The function of the upper and lower splines is to avoid the end points having too much influence over the overall shape of the relationship. It is possible to choose more than two knots but we have not pursued this. We require that the

relationship, as far as possible, be decreasing (or increasing) over the range of class sizes from about 15 to 30. The knots finally chosen are at class sizes of 19 and 25, but the overall results are not markedly sensitive to the precise position of the knots.

Literacy

The following plot (Figure 1) fits a model for post-reception literacy score by class size using (1). Here and elsewhere we plot results for the range of class sizes 15-33; outside that range there are only small numbers of classes and the exact relationship is poorly determined.

(Figure 1 here and Table 2)

We see that there is a decreasing test score with increasing class size, with little apparent change between class sizes of about 18 and 25. The details are given in Table 2. The lack of linearity is important since it indicates that the effect of a given reduction in class size depends on the actual size of class itself. In particular, as we shall discuss in more detail later, there does seem to be a class size above which any reduction has a relatively small effect. We have fitted a multilevel model allowing for between-pupil, between-classroom and between-school variation, where the latter accounts for most of the variation, followed by that at the classroom level. In later models we will introduce random coefficients also. We now carry out an adjustment for the baseline literacy test score using linear and quadratic terms.

(Table 3 and Figure 2 here)

Table 3 and Figure 2 show the same relationship for the adjusted post-reception score, but this time there is a steady decrease of adjusted score with class size. This plot is at the mean of the baseline literacy score. Since figure 2 adjusts for prior achievement, it may be that there is purposive selection of children perceived to have lower prior achievement into smaller classes in the range 18-25 and that the adjustment corrects for this.

Adjusting for baseline literacy score may not adequately remove any association between prior achievement and purposive class size assignment. In the next model, therefore, (Tables 4a and 4b) we have also adjusted for baseline mathematics score as well as a number of factors such as the child's age and term of entry. We have allowed the relationship between the end of reception score and the baseline tests to vary with term of entry. We have also included a compositional variable, namely the average baseline literacy achievement of the pupil's year group in terms of the difference between the pupil's own achievement score and that of the year group. We have also studied the use of the class average rather than the year group average, and obtain similar results. The year group average, where there is more than one class per year group will be a more reliable estimate of the underlying average intake achievement. As mentioned above, we may expect an interaction between class size and disadvantage. We have therefore fitted eligibility for free school meals and its interaction with class size and also divided the sample into three groups according to their baseline literacy score; the lowest 25%, the middle 50% and the highest 25% and fitted the interaction of this categorization and class size. This then allows us to plot separate relationships for these groups. Figure 3 and Table 3 show the results of this model. We have only shown plots for the 3 literacy categories since the free meals categorization produces no further differential effects. Note also that these relationships are adjusted for the other factors, including the basic prior literacy and mathematics scores, so that the differences *between* the lines cannot be interpreted as differences between groups. Likewise the vertical scale is for calibration only since the plots are taken at the zero values of the other factors. For free school meals, however, the vertical distance can be interpreted in terms of a group difference.

(Tables 4a and 4b and Figure 3 here)

What we see is that for the highest and middle achievers at baseline, there is a continuing decrease in achievement with increasing class size, but beyond a class size of about 28 this does not continue for the lowest achieving group. Below a class size of about 25 the gain from a reduction in a class size of 10 is about 0.5 standardised score points for the

lowest achieving group, but in the same range is only about 0.2 points for the other pupils. The free meals plot shows a somewhat steeper relationship for those eligible for free school meals, but an overall advantage in small classes for those not eligible.

(Figure 4 here)

There are a number of other features of this model worth mentioning. One is the relationship with the compositional variable, namely the deviation of the prior achievement from the year group prior achievement mean. Figure 4 is a plot of the adjusted score against this difference, with details in Table 4a. The individual pupil prior literacy score has been adjusted for and it is interesting to note that there is a strong effect the further below the year group mean a pupil is whereas there is relatively little effect as a pupil moves above the mean. There are no significant interactions with the prior literacy score itself. Note also that boys achieve less progress than girls.

(Figure 5 here)

Also of interest for educational policy is the effect of Spring/Summer entry, after adjusting for age. Figure 5 shows that only for the high achievers at entry is there little difference; as the entry achievement decreases so the discrepancy between the Autumn and later entry pupils increases, being 0.6 standardised score points at the mean achievement level.

Mathematics

For mathematics we have followed a similar series of analyses. Figure 6 and Table 6 show the basic relationship for end of reception mathematics by class size. The knots for the curve are chosen to be the same as for literacy, and again the results are robust against changes to these. As with literacy there appears to be only a small amount of change for class sizes from about 20 to 25, with a clearer relationship emerging after adjusting for baseline achievements (Table 6 and Figure 7).

(Table 5 and Figure 6 here)
(Table 6 and Figure 7 here)

Figure 8 and Tables 7a and 7b show the pattern after adjusting for the same set of other factors as for literacy, including baseline literacy achievement. For mathematics there is no interaction with class size. Like literacy, it is the low achievers at baseline who appear to have a larger effect from being in a small class with an increasing reduction in expected achievement up to a class size of about 30. For the middle and high baseline pupils there is little change above a class size of about 22.

(Tables 7a and 7b and Figure 8 here)

Figure 9 shows the relationship between adjusted score and deviation from the class mean baseline mathematics achievement. The pattern is similar to that for literacy. For boys now the coefficient is statistically not significantly different from zero. As with literacy, a Spring or Summer entry is associated with lower adjusted achievement and with all prior achievement levels being associated with a substantially lower achievement; about 0.3 standardised units at the mean achievement level (Figure 10).

(Figures 9 & 10 here)

Discussion

We know from many studies a good deal about the main factors influencing children's progress after entry to school (e.g., Tizard et al, 1988). We know that children's skills and knowledge on entry to school are important determinants of such progress, and we know that income levels (e.g., as indicated by free school meal eligibility), and gender are also important. We also know that home influences and parental input are important, as are endogenous or within-child factors, such as intelligence and ability to concentrate. Over and above these influences the influence of school experiences are bound to be relatively

small, and, as part of that, the influence of class size is bound to be even smaller. Given this, the effect of class size that we find in this study can be seen as impressive. These analyses have demonstrated a clear effect of class size difference on children's academic attainment over the reception year, both before and after adjusting for possible confounding factors. They have also shown some differential effects for the initial low achievers in the case of literacy and for those eligible for free school meals. For mathematics the pattern is somewhat different than for literacy, most noticeably in the small amount of change for class sizes between 25 and 30. As we suggested in the introduction, there remains the possibility that further confounding factors exist which could modify our conclusions. Strictly speaking, results may not generalise to other parts of the U.K. where education policy and practice may vary. These are areas for further research, but our view is that the present study provides the most extensive prima facie evidence for the existence of a real causal effect of class size on achievement,

The average change of post reception mathematics score with age is 0.73 standardised points per year. This is the 'contemporaneous' rate of change and is not the same as the average change that an average pupil will achieve over 1 year. Goldstein and Fogelman (1974) estimate the ratio of the latter to the former, at age 11, to be 2.6 and if this were applied here would imply that a difference of 1 standardised point was equivalent to about 2 years progress. For literacy the corresponding number of years for a difference of 1 standardised score point is estimated to be 1.7 years, resulting in an estimated gender difference in favour of girls of about 4 months. For literacy this would also imply that a decrease of class size of 10, below 25, is associated with a gain of about one years achievement for the lowest achieving group and about 5 months for other pupils. These estimates are impressive, though very rough and should be treated with caution.

There are important policy implications of these results. In general the results support the use of small classes during KS1, or, more precisely, during the first, reception year. However, the results also allow a more specific picture in terms of the children who most benefit. We have found evidence that small classes appear to work best in literacy for children who are most in need academically, that is, those with the lowest school entry

scores who thus have the most ground to make up. For small class sizes there is an advantage to those not eligible for free meals. Both findings, i.e., concerning achievement levels on entry and income levels, suggest where targeting of resources (in this case small classes and those on relatively low incomes) might be best directed.

The picture is more complicated when it comes to judgements about optimum class sizes. For maths and for the lowest attainers in literacy there is a tendency for class size effects to be most marked from 25 and smaller. There are other indications in our study that 25 may be an important number of children, below which relationships with classroom processes, such as the number and size of within class groups, become most evident (Blatchford, Baines, Kutnick and Martin, 2001). This might be contrasted with results from US studies which suggest that class sizes below 20 are crucial for effects to be found; however, as we have seen, studies such as the STAR project are limited in that they do not allow relationships between attainment and class size to be tested across the full range of class sizes.

In future papers we will extend our analysis in a number of ways. Statistical models will be extended to allow for the possible influences of other compositional features of the class, for example, it might be expected that for a given size of class, a child will make more progress if the average ability level of the class is higher, and, conversely, less progress if there are more children who are financially poor, are achieving poorly at pre-test, and have behavioural difficulties. We will also include in models information on classroom processes and behaviour. In other papers we show that class size differences are related to several aspects of classroom processes, including teacher/pupil interactions, pupil attentiveness and peer relations (Blatchford, Edmonds & Martin, submitted for publication), teaching (Blatchford, Moriarty, Edmonds & Martin, in press) and within class grouping practices (Blatchford, Baines, Kutnick & Martin, 2001). A next step is to include information on these and other processes into models predicting progress to examine to what extent they mediate, or explain, the class size effect. In addition, the progress made during years 1 and 2 in relation to class size is being studied and will be reported elsewhere. A preliminary analysis of the effects of extra staff and adults generally has been carried out, but these

variables appear to show no additional effect for either literacy or mathematics, once the other variables in the analysis have been included. Further work on this topic is in progress. Finally, these models will be replicated on the second cohort and a combined analysis will be produced based on larger numbers of classes and pupils.

Figures

Figure 1

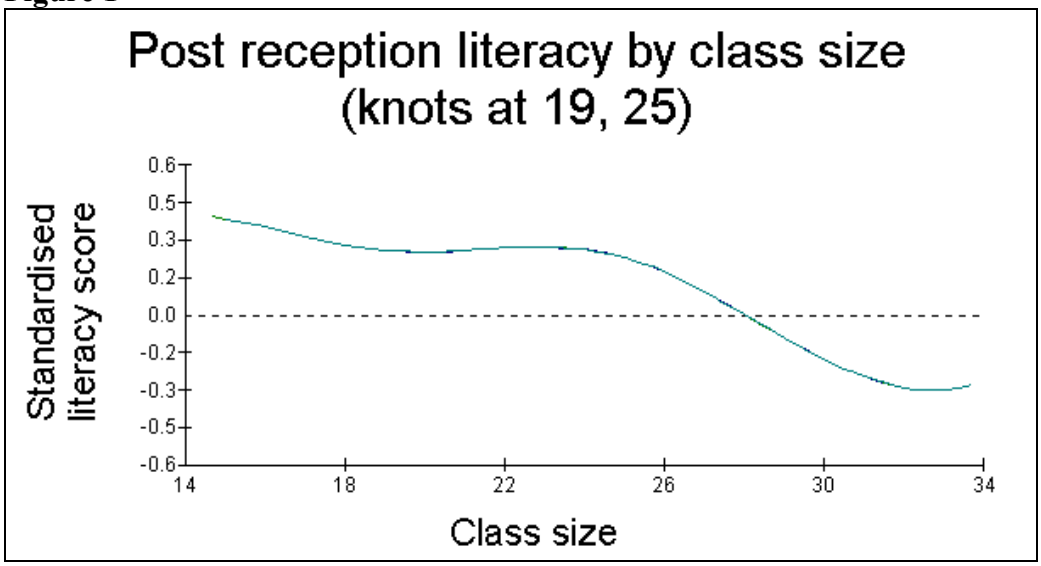


Figure 2

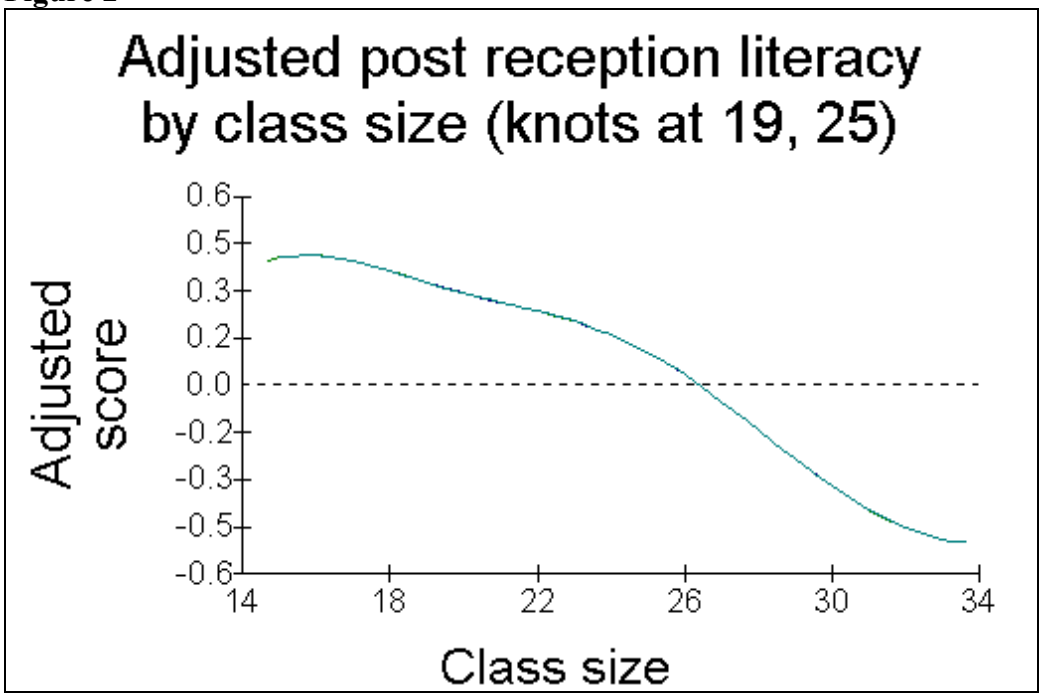


Figure 3

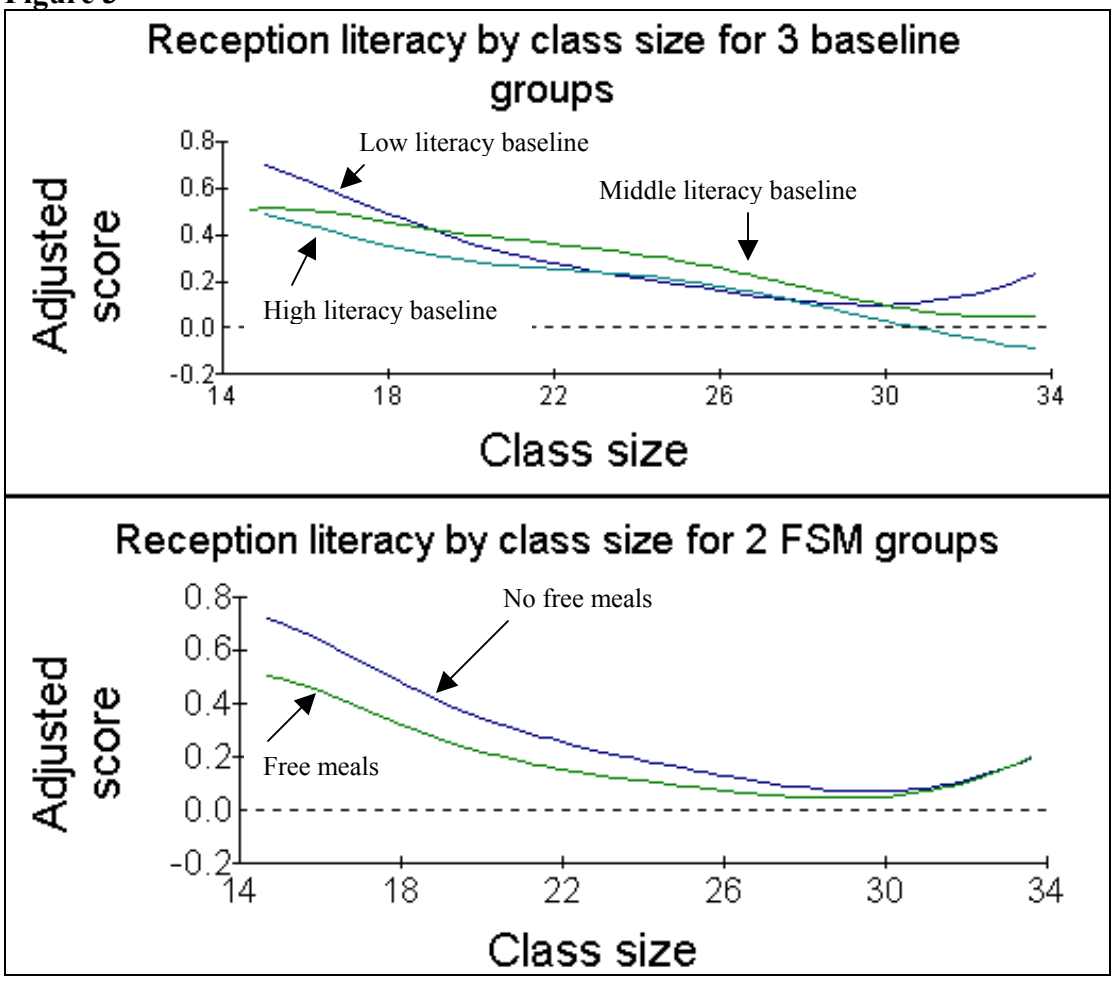
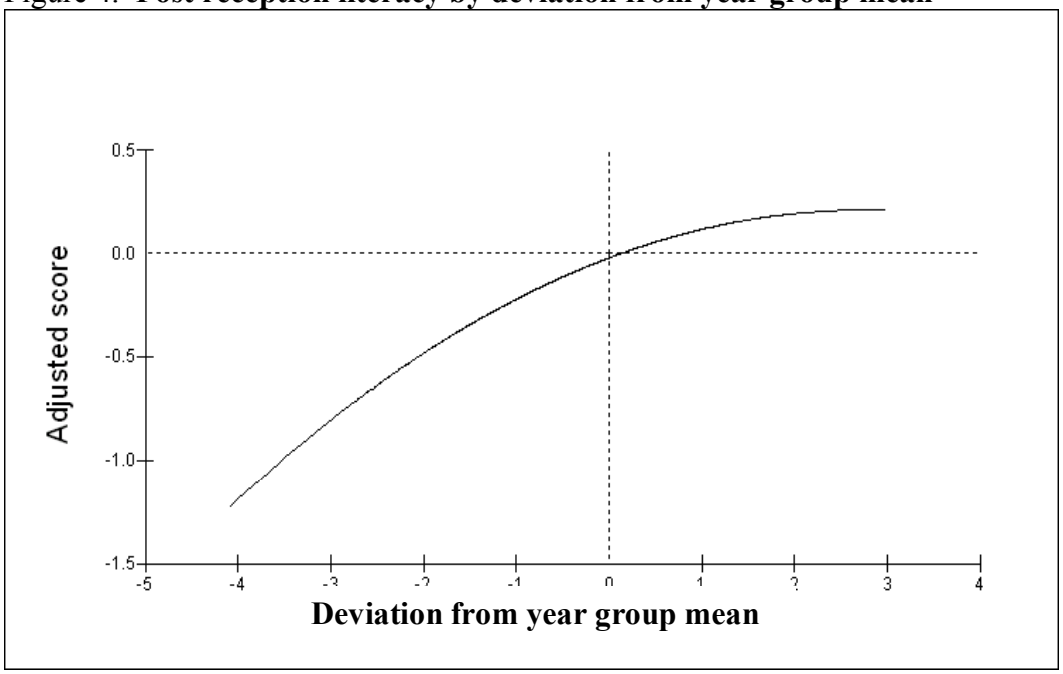
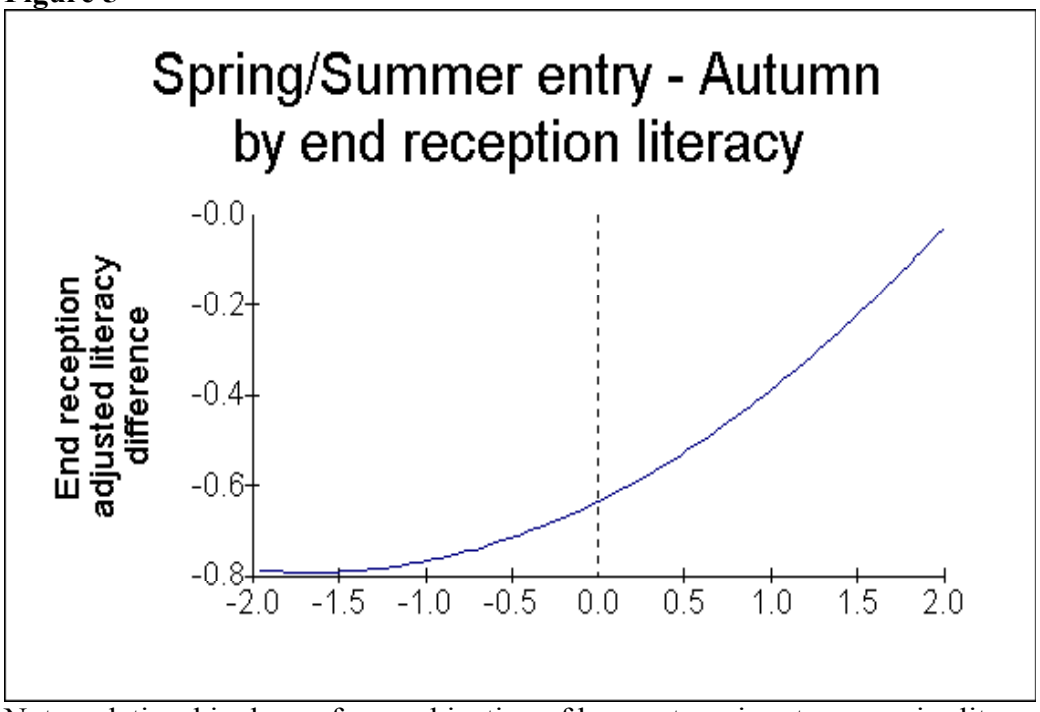

Reception literacy by class size for 3 baseline groups
Adjusted score
0.8
0.6
0.4
0.2
0.0
-0.2
Low literacy baseline
Middle literacy baseline
High literacy baseline
14
18
22
26
30
34
Class size
Reception literacy by class size for 2 FSM groups
Adjusted score
0.8
0.6
0.4
0.2
0.0
-0.2
No free meals
Free meals
14
18
22
26
30
34
Class size

Figure 4. **Post reception literacy by deviation from year group mean**



Note: relationship drawn for combination of base categories at mean prior literacy score.

Figure 5



Note: relationship drawn for combination of base categories at mean prior literacy score.

Figure 6

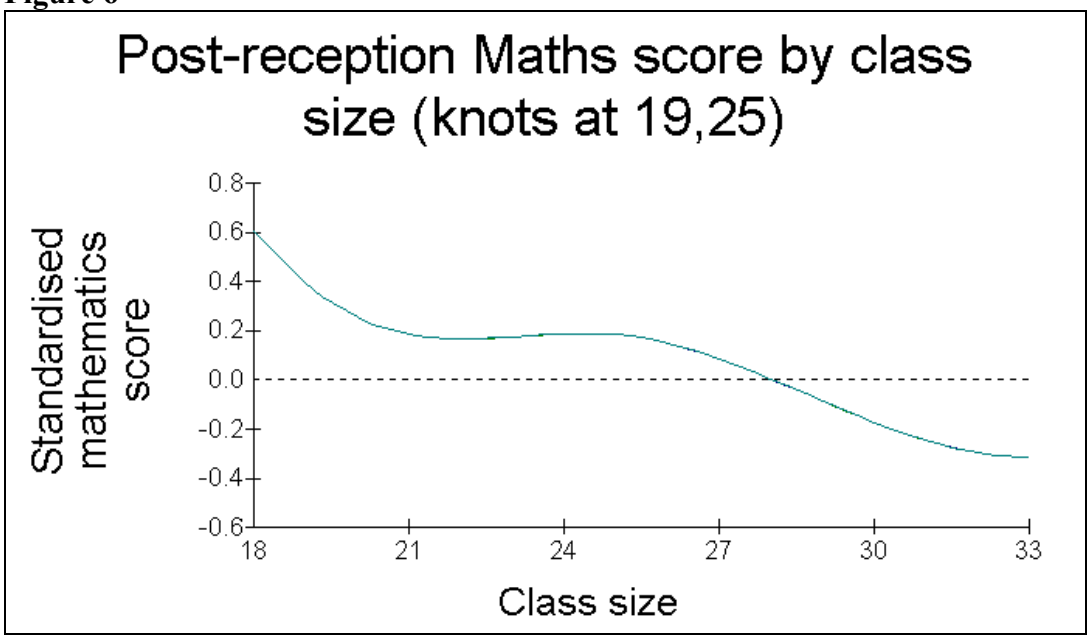


Figure 7

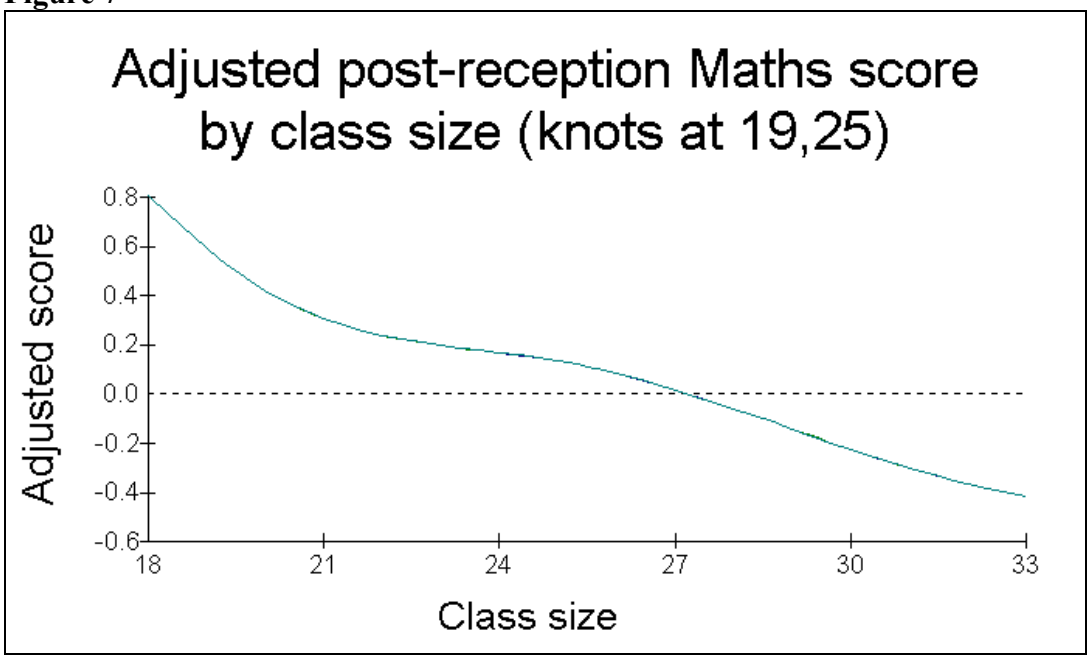


Figure 8

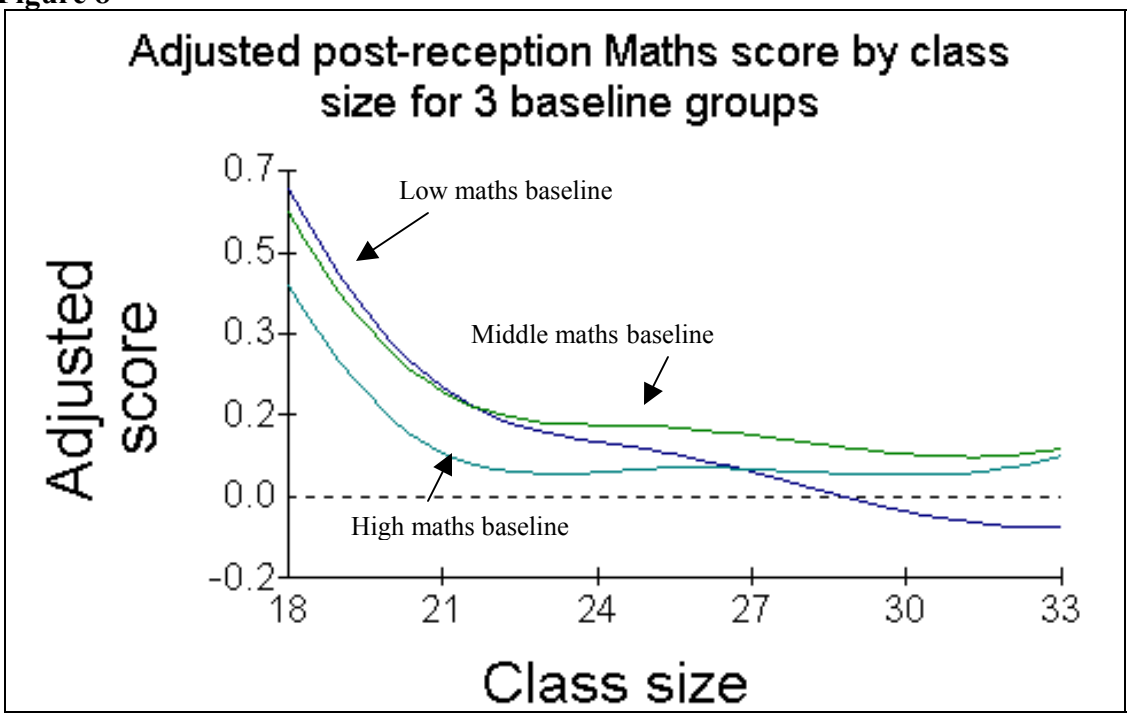
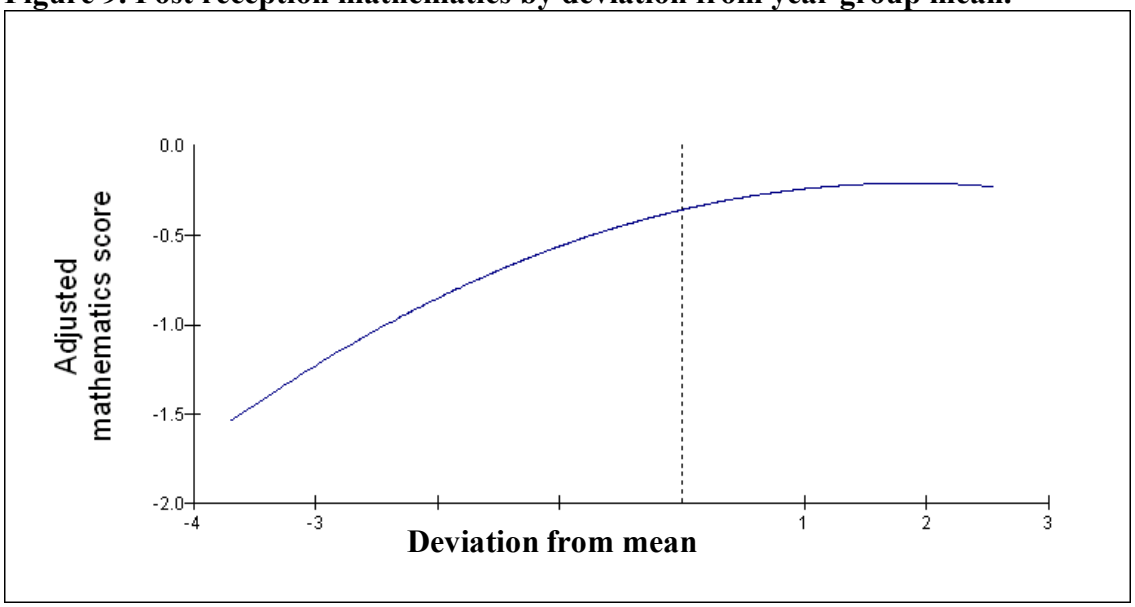
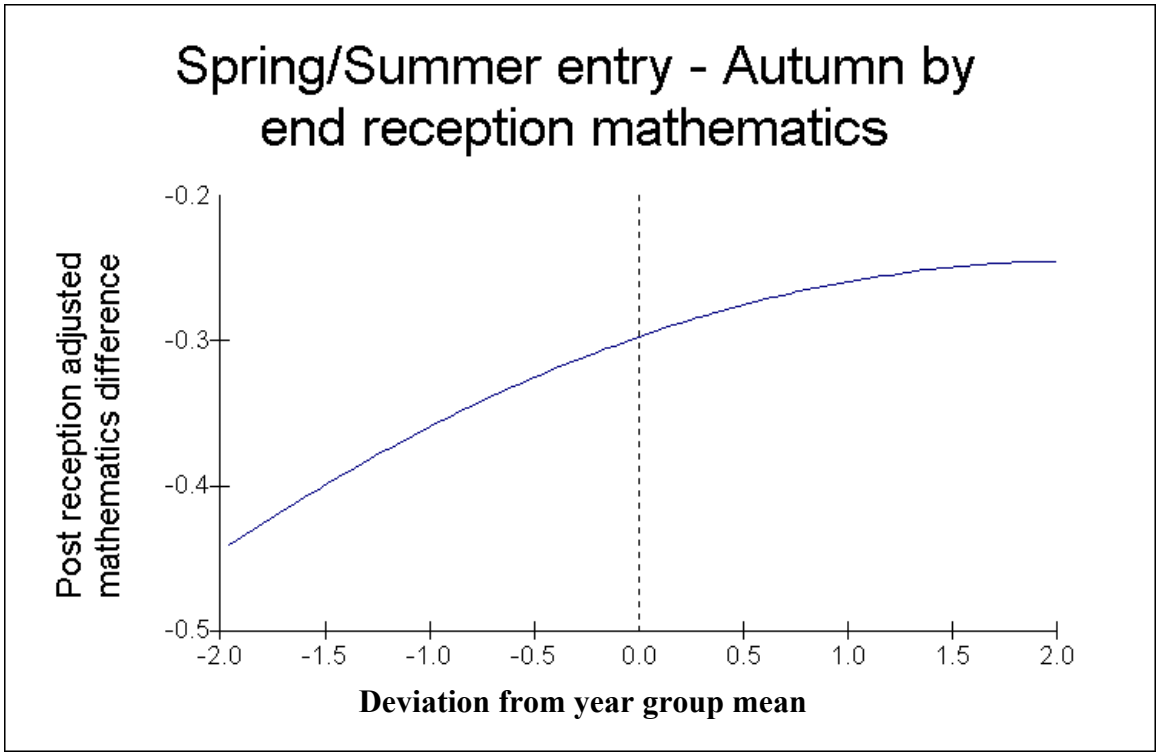


Figure 9. Post reception mathematics by deviation from year group mean.



Note: relationship drawn for combination of base categories at mean prior mathematics score.

Figure 10



Note: relationship drawn for combination of base categories at mean prior mathematics score.

Tables

Table 1. Variables used in analysis	
Variable	**Description**
Average class size	Registered class size experienced by pupil averaged over terms in reception class
Entry time	Whether entered school in Autumn (78%), Summer or Spring
Literacy	Assessment of reading at baseline and end of reception year.
Mathematics	Assessment of mathematics at baseline and end of reception year.
Age	Age at Sept 1 1996
Free school meals	Whether entitled to free school meals

Table 2. Post reception literacy score by class size using model (1). Variance components model

Fixed	Estimate	Standard error
Intercept	-0.54	0.18
Class size	-0.30	0.10
Class size^2	-0.036	0.017
Class size^3	-0.0014	0.0008
Upper cubic spline	0.0029	0.0012
Lower cubic spline	0.0027	0.0022
Random		
Between school variance	0.096	0.027
Between class variance	0.142	0.024
Between pupil variance	0.771	0.015
-2*loglikelihood	15662.3	

Class size is the regular class size averaged over all three terms that a child was in the school. Measured about an origin of 30.
Squared and cubic terms are fitted as indicated.
5870 children used in analysis.

Table 3. Post reception literacy score by class size using model (1) adjusting for baseline literacy score. Variance components model

Fixed	Estimate	Standard error
Intercept	-0.524	
Class size	-0.205	0.090
Class size^2	-0.019	0.015
Class size^3	-0.0007	0.0007
Upper cubic spline	0.0016	0.0011
Lower cubic spline	0.0026	0.0026
Baseline literacy score	0.760	0.011
(Baseline literacy score)^2	0.069	0.007
Random		
Between school variance	0.130	0.028
Between class variance	0.124	0.020
Between pupil variance	0.364	0.007
-2*loglikelihood	10499.5	

Class size is the regular class size averaged over all three terms that a child was in the school. Measured about an origin of 30.
5374 children used in analysis

Table 4a. Post reception literacy score by class size using model (1) adjusting for baseline literacy score, term of entry, free school meals, class average baseline literacy score, mathematics baseline, age and gender. Random coefficients model.

Fixed	Estimate	Standard error
Intercept	-0.011	0.163
Class size (a)	-0.062	0.085
Class size^2 (b)	-0.0064	0.013
Class size^3	-0.0004	0.0006
Upper cubic spline	0.0009	0.0010
Lower cubic spline	0.0016	0.0019
Literacy baseline score (c)	0.336	0.069
(Literacy baseline score)^2 (d)	0.062	0.023
Maths baseline score	0.300	0.021
Literacy baseline score – year group average score	0.171	0.065
(Literacy baseline score – year group average score)^2	-0.034	0.019
Entry in Spring or Summer* (e)	-0.622	0.043
Eligible for free meals* (f)	-0.021	0.034
Boy* (g)	-0.035	0.016
Age in years on Sept 1 of reception year	0.017	0.035
Literacy middle group** (h)	-0.067	0.049
Literacy high group** (i)	-0.070	0.066
Maths middle group** (j)	0.024	0.045
Maths high group** (k)	0.026	0.063
Interactions:		
(a) x (h)	-0.037	0.013
(a) x (i)	-0.044	0.017
(a) x (j)	0.010	0.006
(a) x (k)	0.023	0.008
(b) x (h)	-0.003	0.001
(b) x (i)	-0.006	0.003
(c) x (e)	0.180	0.027
(d) x (e)	0.054	0.021
(a) x (f)	0.008	0.011
(b) x (f)	-0.0003	0.0010

Class size is the regular class size averaged over all three terms that a child was in the school. Measured about an origin of 30.
4833 children used in analysis
Variables marked * are dummy (0,1) variables where the variable name indicates the category coded 1.
Variables marked ** are categories for baseline scores such that lowest category (the base category) is the lowest 25%, the middle is the next 50% and the high is the top 25%

Likelihood ratio tests:

Interaction with term of entry; $\chi^2 = 45.7$ (2 d.f.), P<<0.001

Interaction with eligibility for free school meals; $\chi^2 = 3.6$ (2 d.f.), P=0.17

Interaction with Literacy and Maths groups; $\chi^2 =$ (6 d.f.), P=0.003

Significance tests for variables not involved in interactions can be obtained from the estimated standard errors.

Table 4b. Random parameters. Covariance matrices (s.e. in brackets)

Between-schools	Intercept	Class size	Literacy baseline	(Literacy baseline)^2
Intercept	0.15 (0.031)			
Class size	0.006 (0.003)	0.0010 (0.0004)		
Literacy baseline	-0.045 (0.009)	0.0002 (0.0011)	0.0224 (0.0045)	
(Literacy baseline)^2	-0.024 (0.006)	-0.0010 (0.0007)	-0.0002 (0.0020)	0.0061 (0.0018)

Between class variance	0.063 (0.011)
Between pupil variance	0.290 (0.006)
-2 * loglikelihood	8450.6

Likelihood ratio tests:

Coefficient of square of literacy baseline score; $\chi^2 = 54.0$ (4 d.f.), P<<0.001

Coefficient of class size; $\chi^2 = 11.1$ (4 d.f.), P=0.03

Table 5. Post reception mathematics score by class size using model (1). Variance components model

Fixed	Estimate	Standard error
Intercept	-0.840	
Class size	-0.483	0.161
Class size^2	-0.074	0.026
Class size^3	-0.004	0.001
Upper cubic spline	0.0053	0.0021
Lower cubic spline	0.0277	0.0013
Random		
Between school variance	0.065	0.026
Between class variance	0.161	0.027
Between pupil variance	0.764	0.015
-2*loglikelihood	14775.8	

Class size is the regular class size averaged over all three terms that a child was in the school. Measured about an origin of 30.
5560 children used in analysis.

Table 6. Post reception mathematics score by class size, adjusting for baseline mathematics and literacy scores. Variance components model

Fixed	Estimate	Standard error
Intercept	-0.675	
Class size	-0.350	0.134
Class size^2	-0.051	0.022
Class size^3	-0.0027	0.0010
Upper cubic spline	0.0036	0.0017
Lower cubic spline	0.0323	0,0087
Baseline literacy score	0.352	0.017
(Baseline literacy score)^2	-0.005	0.009
Baseline maths score	0.379	0.017
(Baseline maths score)^2	0.012	0.009
Random		
Between school variance	0.097	0.026
Between class variance	0.136	0.022
Between pupil variance	0.384	0.008
-2*loglikelihood	10289.3	

Class size is the regular class size averaged over all three terms that a child was in the school. Measured about an origin of 30.
5143 children used in analysis

Table 7a. Post reception mathematics score by class size using model (1) adjusting for baseline literacy score, term of entry, free school meals, class average baseline literacy score, mathematics baseline, age and gender. Random coefficients model.

Fixed	Estimate	Standard error
Intercept	-0.320	
Class size (a)	-0.197	0.130
Class size^2 (b)	-0.032	0.021
Class size^3	-0.0019	0.0010
Upper cubic spline	0.0023	0.0017
Lower cubic spline	0.0192	0.0080
Maths baseline score (c)	0.072	0.071
(Maths baseline score)^2 (d)	0.072	0.023
Literacy baseline score	0.342	0.017
(Literacy baseline score)^2	-0.010	0.009
Maths baseline score – year group average score	0.288	0.067
(Maths baseline score – year group average score)^2	-0.065	0.018
Entry in Spring or Summer* (e)	-0.525	0.048
Eligible for free meals*	-0.086	0.028
Boy*	-0.005	0.018
Age in years on Sept 1 of reception year (centered at 4.0)	0.134	0.040
Maths middle group** (j)	0.123	0.053
Maths high group** (k)	0.079	0.072
Interactions:		
(a) x (j)	0.015	0.008
(a) x (k)	0.024	0.009
(c) x (e)	0.095	0.030
(d) x (e)	0.0002	0.023

Class size is the regular class size averaged over all three terms that a child was in the school. Measured about an origin of 30.
4691 children used in analysis
Variables marked * are dummy (0,1) variables where the variable name indicates the category coded 1.
Variables marked ** are categories for baseline scores such that lowest category (the base category) is the lowest 25%, the middle is the next 50% and the high is the top 25%

Likelihood ratio tests:

Interaction with term of entry; $\chi^2 = 11.0$ (2 d.f.), P=0.004

Interaction with Maths groups; $\chi^2 = 6.2$ (2 d.f.), P=0.05

Significance tests for variables not involved in interactions can be obtained from the estimated standard errors.

Table 7b. Random parameters. Covariance matrices (s.e. in brackets)

Between-schools	Intercept	Maths baseline	(Maths baseline)^2
Intercept	0.083 (0.023)		
Maths baseline	-0.017 (0.006)	0.008 (0.003)	
(Maths baseline)^2	-0.016 (0.005)	0.002 (0.002)	0.007 (0.002)
Between class variance		0.100 (0.017)	
Between variance		0.359 (0.008)	
-2 * loglikelihood		9141.3	

Likelihood ratio test:

Coefficient of square of Maths baseline score; χ^2 =36.0 (3 d.f.), P<<0.001

References

Avon Reception Entry Assessment (1996) Avon Education Department and the Institute of Education

Blatchford, P., Edmonds, S. & Martin, C. Class size, pupil attentiveness and peer relations (submitted)

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*, 20, 4, 411-428

Blatchford, P., Baines, E., Kutnick, P., & Martin, C. (2001) Classroom contexts: connections between class size and within class grouping. *British Journal of Educational Psychology*, 71, 283-302

Blatchford, P., Goldstein, H. and Mortimore, P. (1998) Research on class size effects: a critique of methods and a way forward. *International Journal of Educational Research*, 29, 691-710

Blatchford, P., Moriarty, V., Edmonds, S., and Martin, C. (in press) Relationships between class size and teaching: a multi-method analysis of English infant schools. *American Educational Research Journal*

Finn, J.D. and Achilles, C.M. (1999) Tennessee's Class Size Study: findings, implications, misconceptions. *Educational Evaluation and Policy Analysis*, 21, 2, 97-109.

Goldstein, H. and Blatchford, P. (1998). Class size and educational achievement: a review of methodology with particular reference to study design. *British Educational Research Journal* 24: 255-268.

Goldstein, H. and Fogelman, K. (1974). Age Standardisation & Seasonal Effects in Mental Testing. *British Journal of Mathematical and Statistical Psychology* 44: 109-115.

Hastie, T. J. and Tibshirani, R. J. (1990). *Generalized Linear Models*. London, Chapman and Hall:

Molnar, A., Smith, P., Zahorik, J., Palmer, A., Halbach, A., and Ehrle, K. (1999) Evaluating the SAGE program: a pilot program in targeted pupil-teacher reduction in Wisconsin. *Educational Evaluation and Policy Analysis,* 21, 2, 165-177.

Reading Progress Tests: (2000) Literacy Baseline, London: Hodder & Stoughton

Slavin, R. (1989). Class size and student achievement: small effects of small classes. *Educational psychologist* 24: 99-110.

Tizard, B., Blatchford, P., Burke, J., Farquhar, C. and Plewis, I. (1988) *Young Children At School In The Inner City.* Hove: LEA.

Yang, M., Goldstein, H., Omar, R., Turner, R., and Thompson, S. G. (2000). Meta analysis using multilevel models with an application to the study of class size effects. *Journal of the Royal Statistical Society, Series C* 49: 1-14.

Word, E. R., Johnston, J., Bain, H. P., Fulton, B. D., et al. (1990). *The state of Tennessee's student/teacher achievement ratio (STAR) project: Technical report 1985-90.* Nashville, Tennessee State University.