Economica (2003) **70**, 31–53

Do Peer Groups Matter? Peer Group versus Schooling Effects on Academic Attainment

By DONALD ROBERTSON† and JAMES SYMONS‡

† *University of Cambridge* ‡ *University College London*

Final version received 23 July 2001.

This paper estimates an educational production function. Educational attainment is a function of peer group, parental input and schooling. Conventional measures of school quality are not good predictors for academic attainment, once we control for peer group effects; parental qualities also have strong effects on academic attainment. This academic attainment is a then a key determinant of subsequent labour market success, as measured by earnings. The main methodological innovation in this paper is the nomination of a set of instruments, very broad regions of birth, which, as a whole, pass close scrutiny for validity and permit unbiased estimation of the production function.

I don't understand all this fuss about education. None of the Paget family can read or write, but they do very well.

William Lamb, 2nd Viscount Melbourne

INTRODUCTION

This paper describes an empirical investigation of the causes of academic achievement in British children at age 11. We investigate as well the relationship between such achievement and subsequent labour market success.

There is a view in the literature of the economics of education that, to a first approximation at least, the academic attainment of children in schools is completely explained by the education, income and social class of their parents, and by the levels of these factors in the child's peer group; school quality as measured by conventional inputs such as class size, teacher experience and general school expenditure are typically found to have only minor effects. This view originates with the work of the Coleman report (Coleman *et al*. 1966), and is extensively reviewed and discussed in Hanushek (1986). The Coleman report emphasized the pre-eminence of peer group effects on educational attainment. More recent studies are Summers and Wolfe (1977), Henderson *et al*. (1978), Borjas (1992), Hoxby (2000), Sacerdote (2000) and Epple *et al*. (2000). One expects peers to affect achievement directly (e.g. helping each other with course work) and also via values, though it seems to us likely that the latter effect predominates.

Recently Card and Krueger (1992) have challenged this accepted wisdom. They find a significant correlation between the returns to schooling in the United States and state-wide educational quality as measured by conventional inputs. Whereas this study gives results for labour market outcomes, and the Coleman–Hanushek evidence relates to achievement, the results are clearly at variance with the earlier literature. The question is of immense significance for public policy: if Card and Krueger are right, the problem of the poor could be alleviated within one or two generations by increased school expenditures, while if they are wrong such expenditures would be a complete waste.

We study these issues using a cohort of British children born in a particular week in 1958 (the National Child Development Study, NCDS). Section I describes the model we shall use to address these issues. Section II describes the estimation results and econometric issues in greater detail and briefly discusses the problems raised by attrition in the NCDS. Section III provides further analysis of the peer group effects, and Section IV concludes.

The paper presents estimates of an educational production function to distinguish the effects of peer groups and conventional inputs. The advance we make is in our treatment of the potential endogeneity of these inputs. We propose as instruments indicator variables for ten regions of birth in the United Kingdom. These instruments are open to the objection that they are endogenous if parents actively choose location with a view to enhancing the attainment of their children. We conclude however that such effects are unimportant in our data. First, there does not appear to be much evidence of parental moves to regions of good schools following the birth of first children, which should be observable if an important proportion of parents actively choose regions in this way. Second, the residuals from our models are found to be adequately orthogonal to the instruments via the conventional Sargan test.

Another innovation in this study is careful analysis of the properties of our instrumental variables estimators. Recent econometric literature has highlighted problems for inference arising when instruments are not particularly strong; i.e. they do not separately explain a great deal of the endogenous variables. Our instruments are neither very strong nor very weak, and to resolve the resulting ambiguity we conducted Monte Carlo simulations of our model. We found that, while bias is a minor problem in our data, its direction tends to understate the true magnitude of our key peer group variable.

We find in fact that the IV estimates of the peer group effect are larger in magnitude than OLS, suggesting that measurement error bias dominates choice-theoretic bias effects. Active choice of peer group does not appear to cause substantial problems for our estimates.

All in all, we believe the calibre of these new estimates of the peer group effects is higher than that in the existing literature. Our evidence points towards the parents and peers theory of educational attainment in children, giving only a minor role to school inputs. This attainment is then rewarded in the labour market.

I. The Educational Production Function

The central theoretical concept in this study is the educational production function. Attainment is viewed as being generated by peer group inputs, parental inputs and conventional school inputs. There are a number of possible channels through which the peer group can influence attainment. A better peer group may be comprised of better behaved children, who may also value education more highly, and whose parents become more involved in the educational process. The consequence of this is a better educational environment for all children in the class, irrespective of their specific origin.

In principle, attainment at any date will depend on the entire history of inputs prior to that date. Attainment will also depend on inherent ability, some part of which may be genetic (and thus could be thought of as an original parental input), and some part of which may be due to chance.

Following Hanushek (1992), we write attainment at time t, A_t, to be some function

$$A_t = \Phi(\mathbf{P}_t, \mathbf{F}_t, \mathbf{S}_t), \tag{1}$$

where $\mathbf{P}_t, \mathbf{F}_t$ and $\mathbf{S}_t$ are the streams of past peer group, family and schooling inputs, respectively, obtained by time t. If prior achievement is also observed, then, under certain assumptions, this function can be rewritten as

$$A_t = \phi(A_{t-1}, P_t, F_t, S_t), \tag{2}$$

where P_t, F_t and S_t now represent the inputs from $t-1$ to t. Parents have an intertemporal utility function defined over the ultimate welfare (which we assume here is indexed by attainment) of their children, together with their own consumption and leisure. This utility function is maximized subject to both financial and time constraints and the production function. This makes clear, as is almost always the case in production theory, that the observed factor inputs are endogenous. The problem that will usually arise in estimating (2) is that parents who are exceptionally ambitious for their children (in a manner unobserved by the econometrician) will choose high levels of the productive inputs as well as aiding the child directly. This gives biases if simple regression estimates are interpreted as structural parameters of the production function. These biases apply in principle to most of the variables we investigate below, from those measuring school quality to family structure variables such as number of children or presence of a father. The usual solution is to apply an instrumental variable estimator. The outcome from the optimization above would give reduced forms of the endogenous variables as functions of exogenous variables, most notably prices of labour and prices of schooling inputs. These we do not observe directly, but, if we assume that these prices are constant across local markets, then regional indicator variables will then be available as potential instruments, provided that preferences and technology do not vary systematically over regions, and that an individual's region is not itself endogenous. Econometric tests allow us to investigate the validity of these variables as instruments in the current context, and we shall use these instruments to correct endogeneity bias where it seems most likely (the variables most easily chosen by parents), and in variables of particular interest to this study, i.e. those reflecting school and peer group. In general we shall treat other variables, such as mother's labour force status as exogenous. Where bias does occur, its likely direction is often known, thus aiding inference.

II. The Data and Empirical Estimation: the National Child Development Study

The National Child Development Study is an ongoing survey of a particular cohort of the UK population: every child born in the UK during one week in March 1958. Information has so far been collected at five points in their lives on a wide variety of personal, parental, educational and occupational subjects. The first sweep of the survey (NCDS1) collected information in 1965 when the children were 7 years old. This information comprised mainly parental data, a large amount of health information on the child and a certain amount of early schooling data. Subsequent sweeps took place in 1969 (NCDS2), 1974 (NCDS3), 1981 (NCDS4) and 1991 (NCDS5). At each sweep further parental, health and educational data were recorded, including scores on ability tests. At the final two sweeps a large amount of labour market data were also collected, in addition to educational attainment data. At the later sweeps, data from individuals of the same cohort not born in the UK, but now resident in the UK, were also collected. The original NCDS sample consisted of 18,359 individuals. It may be viewed as a representative subsample of the population of the UK born in the late 1950s.

Unfortunately, not only has there been attrition over the life of the survey, but not all variables are available for all individuals remaining in the study. Consequently each estimation is performed over the largest possible subsets for whom the required data are available. If attrition and non-response are uncorrelated with the explanatory variables in the regressions, then our estimated coefficients remain unbiased. Below we perform some simple analyses regarding the non-responding individuals that will bear on this issue.

Estimation and results

The first set of reported regressions link measured abilities of the individual at age 7 to parental and family variables. Ability is measured by test scores of maths and reading attainment. Figures 1 and 2 graph the distribution of the percentage raw scores for the sample. An immediate problem is the bunching

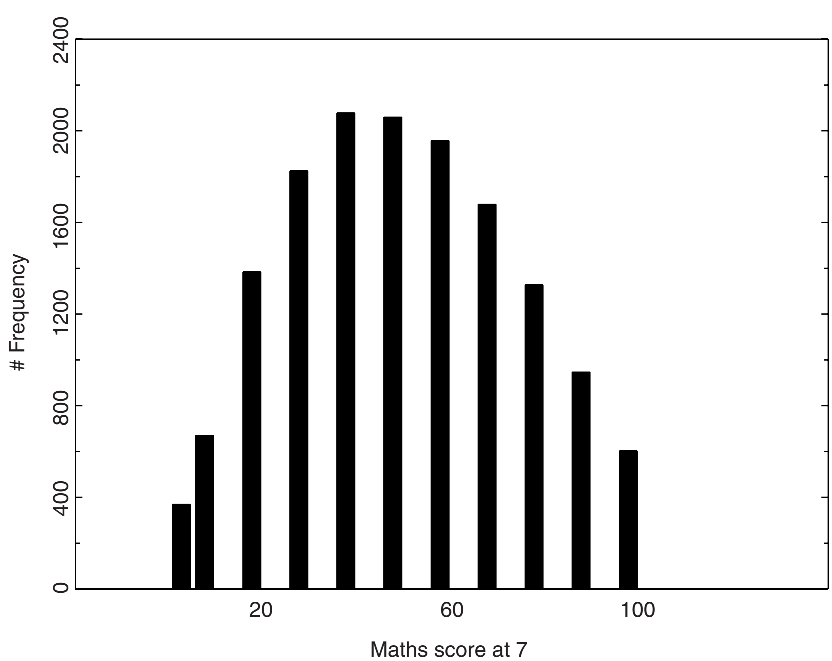

FIGURE 1

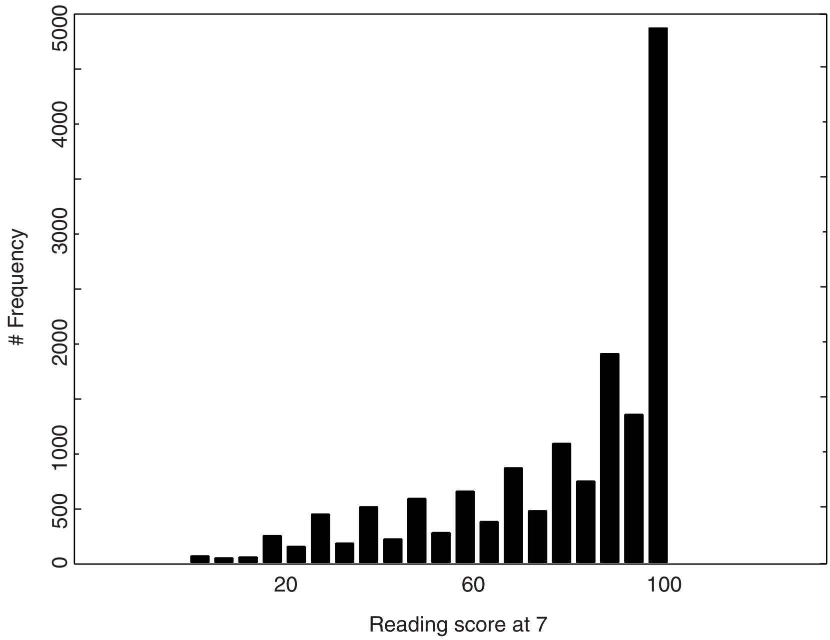

FIGURE 2

of the reading scores at the upper limit. It seems clear that the reading test here is too coarse and is failing to discriminate adequately for the individuals in the upper tail of the distribution of the reading tests. This bunching of scores would clearly be detrimental to our subsequent analysis. The data-set contains a set of variables describing teachers' assessments of the abilities of the children at age 7 in a range of categories/subjects, and Table 1 reports these assessments as explanatory variables in a Tobit regression for reading score, assuming truncation at score 100. The regression results indicate that the explanatory variables have strong predictive power for the reading score. The

TABLE 1
TOBIT REGRESSIONS FOR READING SCORES AT AGE 7

	Coefficient	t-statistic
Oral ability	−1.1	−4.8
Awareness	−1.8	−6.6
Reading ability	−12.4	−45.9
Creativity	−0.6	−2.4
Maths ability	−1.9	−8.4
Comprehension	−6.8	−47.8
Constant	156.7	233.2

Pseudo $R^2 = 0.12$.
Notes:
11,955 uncensored observations; 2797 right-censored observations.
Dependent variable is reading score at age 7, scale 0–100, assumed truncated at 100.
The explanatory variables in this regression are teacher's assessments of child's ability on a scale 1 (highest) to 5 (lowest), except for comprehension marked 1 (highest) to 7 (lowest).

predicted values from this regression are used in place of the raw reading score throughout the subsequent analysis. Figure 3 graphs these predicted values as percentages. Note that the Tobit-corrected predicted values produce a bell-shaped distribution of reading scores.

Our interpretation of the reading and maths scores at age 7 is that they give some measure of initial ability, and possibly some early schooling effects. This initial ability will be a combination of genetic inheritance and pre-school parental inputs which we shall not seek to disentangle. Table 2 relates these maths and (both uncorrected and Tobit-corrected) reading scores at age 7 to a set of parental and schooling variables.

The parental variables may loosely be grouped into variables reflecting parental quality and those reflecting parental time. We proxy parental time by whether the mother works (full time or part time), the presence of a father and the number of children in the household. The quality of parental time is measured by indicators of parental education and social class. We imagine that these quality variables affect attainment both directly (e.g. an educated parent is better able to help with homework) and indirectly via the transmission of values (e.g. punctuality, self-control, ambition). We find that attainment is, in general, negatively related to the number of children in the household (at the time of the tests). Having a working mother reduces attainment at age 7 (particularly if she works full time). It is problematic to interpret these results as the *ceteris paribus* effect of a woman's going out to work, since labour market status is a decision variable of the household. Several different biases are possible. It may be that women who choose to work have a lower level of interest in their children's development. On the other hand, those women who work may be those with higher marginal product, possibly reflected in superior

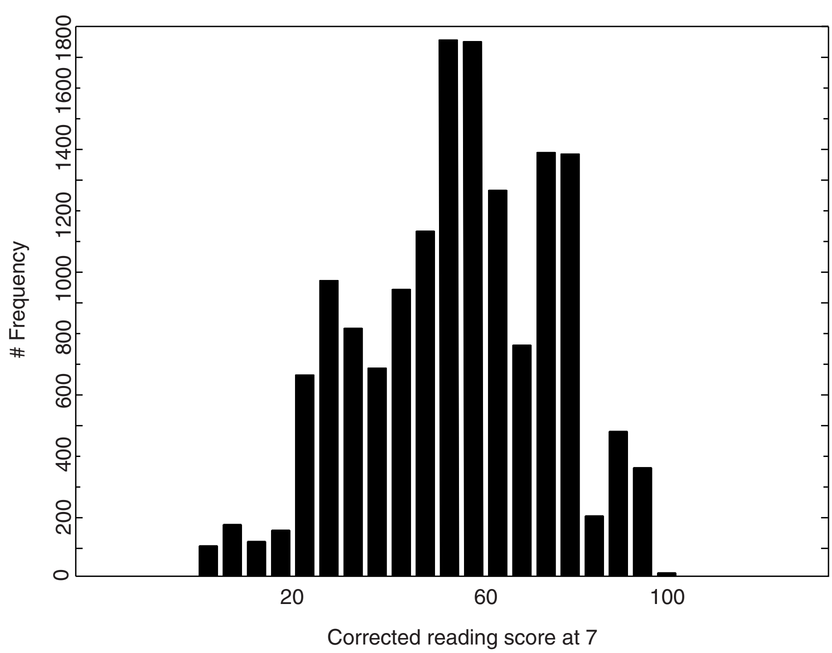

FIGURE 3

TABLE 2
READING AND MATHS SCORES AT AGE 7 (OLS ESTIMATES)

	Reading		Corrected reading		Maths	
Parent quality						
Top SES father	7.8	(10.0)	7.3	(11.2)	6.5	(7.4)
Middle SES father	4.5	(8.1)	3.3	(7.0)	2.7	(4.4)
Father stayed on	5.8	(9.4)	6.2	(12.1)	4.7	(6.8)
Mother stayed on	5.2	(9.2)	5.9	(12.3)	5.8	(9.0)
Parent time						
Mother works FT	−3.0	(3.7)	−1.9	(2.8)	−2.9	(3.1)
Mother works PT	−0.8	(1.4)	−0.6	(1.3)	−0.8	(1.3)
No father	−5.2	(3.7)	−3.9	(3.4)	−1.7	(1.0)
Family size	−2.8	(19.3)	−2.3	(19.1)	−0.8	(5.0)
Sex (=1 if girl)	6.3	(14.0)	5.1	(13.7)	−2.5	(5.0)
Scotland	8.4	(12.7)	5.1	(9.3)	−2.2	(2.9)
Constant	77.5	(103.0)	53.2	(85.0)	51.5	(60.8)
R^2		0.14		0.17		0.054
No. of observations		8749		8701		8733

Notes:
Absolute *t*-statistics in parentheses.
The top SES group is defined as professional, managerial and skilled non-manual occupations; the middle SES group as semi-skilled and skilled manual occupations.
'Stayed on' means remained at school beyond compulsory schooling age.
Family size is the number of children under 21 in the family.
'Mother works FT and PT' are, respectively, dummy variables on whether mother worked full time or part time before the child started school.

child raising. There are also possible income effects. Absence of father has in general a negative effect on attainment, but is statistically significant only for reading. Social class and education of parents typically have the expected effect: the children of the educated and the higher social classes do better. Note that having parents who remained in education beyond the statutory minimum leaving age contributes about ten extra marks (out of 100) in both reading and maths, as does membership of the top socioeconomic status (SES) group (proxied by father's occupation).

We find that girls are better at reading at 7 but worse at maths. Why this should be so is beyond the scope of this paper, but we note that this has been found by other scholars in this area: for example, Harbison and Hanushek (1992) find similar results for rural Brazil. We have experimented also with separate regressions by sex, but found no important differences, except for the constant.[1] It is interesting that we are able to explain less than half the variance in maths scores that we explain for reading scores. It may be that reading skills are easier to impart than mathematical skills. We include a dummy variable for Scotland because the education system there is under a different authority and differs from the rest of the UK in a number of respects (see Feinstein *et al.* 1999). Scottish children at 7 seem more advanced at reading but less advanced at mathematics.

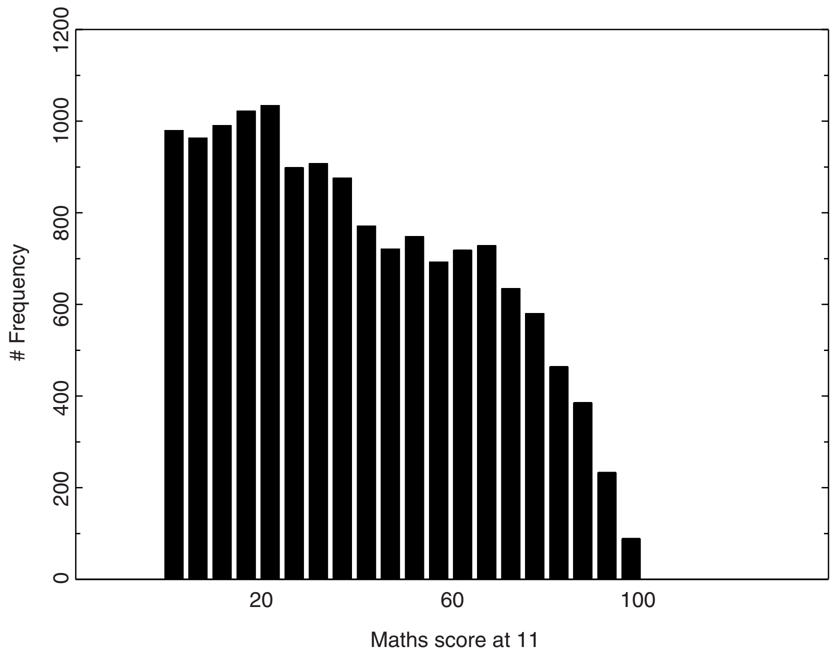

FIGURE 4

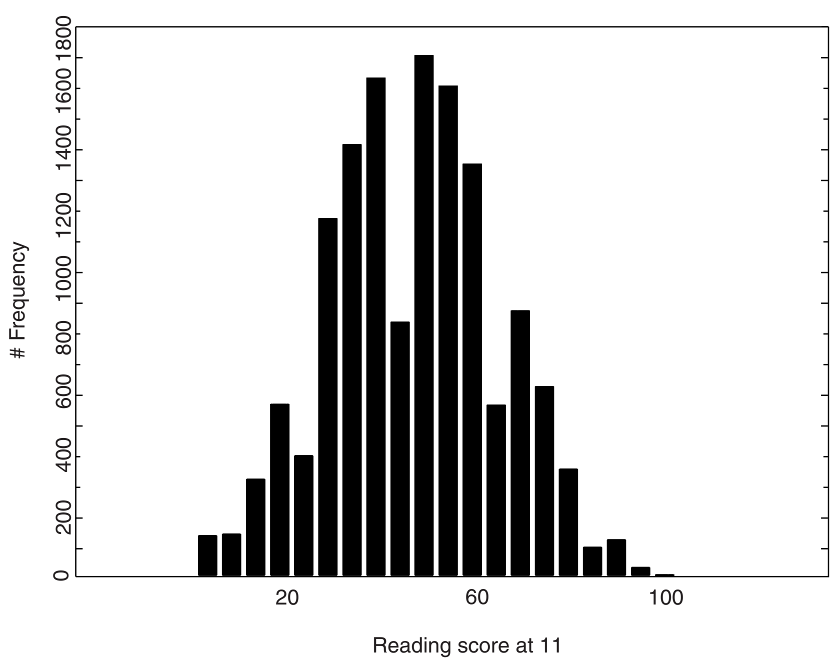

FIGURE 5

Reading and maths tests were re-administered at age 11. The percentage scores are graphed in Figures 4 and 5. Our regressions explain the improvement in scores by this age. Initially, in Tables 3 and 4 we report OLS estimates of the educational production function. We measure the influence of the peer group by the school's report of the percentage of a child's classmates (who typically will not be NCDS children) who have fathers in the top socioeconomic group. There can be a bias towards finding peer group effects in empirical studies if membership of a peer group is endogenous.[2] If parents actively choose the schools their children attend, then one presumes that ambitious parents will seek good schools for their children as well as helping them in other ways. Given that parental ambition is unobserved, empirical studies will tend mistakenly to allocate to the peer group the effect of the unobserved extra parental influence, giving an upward bias to the parameter estimate. A similar selection bias has been studied by Evans *et al.* (1992) for the incidence of teenage pregnancy and school dropout behaviour.

TABLE 3
MATHS IMPROVEMENT, 7–11 YEARS

	Streamed school		Unstreamed school	
Maths score at 7	−0.7	(36.7)	−0.5	(35.8)
Peer group				
Top stream	20.2	(19.5)	—	
Bottom stream	−14.5	(12.8)	—	
% top SES	0.12	(6.1)	0.15	(8.7)
Parent quality				
Father in top SES	3.0	(1.8)	8.1	(6.4)
Father in middle SES	1.2	(1.1)	3.9	(4.7)
Father stayed on	2.8	(2.4)	4.2	(4.4)
Mother stayed on	3.1	(2.9)	5.9	(6.6)
Parent time				
Mother works FT	1.0	(0.8)	−0.2	(0.2)
Mother works PT	2.7	(2.2)	1.4	(1.3)
No father	−2.0	(0.9)	−2.2	(1.1)
Family size	−0.7	(2.3)	−2.1	(8.8)
School quality				
Class size	0.03	(0.4)	0.08	(1.7)
Sex (=1 if girl)	−3.2	(3.8)	−0.1	(0.1)
Scotland	1.8	(1.0)	7.2	(7.8)
Constant	18.2	(5.6)	9.3	(4.0)
R^2	0.54		0.33	
No. of observations	1526		3322	

Notes:
Absolute *t*-statistics in parentheses.
The variable names are largely as in Table 2. Exceptions are that here 'Mother works FT and PT' refer to mother working full time or part time when child is age 11. '% top SES' is the percentage of the child's class with fathers in the top SES class, measured 0–100%, where the definitions of SES are as in Table 2.
Estimation by OLS.

TABLE 4

READING IMPROVEMENT, 7–11 YEARS

	Streamed school		Unstreamed school	
Reading score at 7	−0.7	(31.1)	−0.5	(37.9)
Peer group				
Top stream	8.2	(10.3)	—	
Bottom stream	−8.0	(9.2)	—	
% top SES	0.10	(7.0)	0.06	(5.0)
Parent quality				
Father in top SES	0.1	(0.1)	4.4	(5.5)
Father in middle SES	0.1	(0.1)	1.5	(2.9)
Father stayed on	1.1	(1.3)	1.5	(2.3)
Mother stayed on	1.9	(2.4)	2.5	(4.4)
Parent time				
Mother works FT	0.5	(0.6)	−0.1	(0.2)
Mother works PT	1.0	(1.1)	0.1	(0.1)
No father	0.3	(0.2)	−1.5	(1.1)
Family size	−0.5	(2.2)	−1.0	(6.7)
School quality				
Class size	0.04	(0.8)	0.06	(1.9)
Sex (=1 if girl)	−2.5	(4.1)	−3.0	(6.7)
Scotland	−0.8	(0.6)	−1.4	(2.3)
Constant	24.8	(9.8)	16.7	(10.9)
R^2	0.41		0.34	
No. of observations	1525		3318	

Notes: as for Table 3.

In these data we are also able to investigate a further peer group effect that arises from streaming, the practice of sorting children within schools into ability groups for teaching. About one-third of the children in our sample attended streamed schools, and we report results for them separately in Tables 3 and 4. The NCDS records three ability groups within streamed schools, and each of these will have a potential peer group effect on children placed in them. The information about the stream to which the child belongs is taken at age 11 and there is thus another endogeneity problem, in that children who do exceptionally well between 7 and 11 will tend to be placed in the top stream. Thus, one expects OLS estimates to be biased upwards for the top-stream dummy, and downwards for the bottom-stream dummy.

For each improvement regression, we include the test score at age 7. Our interpretation is that this measures initial ability and may also capture dynamic elements of human capital formation. This variable enables us to judge the extent to which differences in measured ability at age 7 persist to age 11. Since test scores measure ability with error, there is a likely downward bias in the OLS estimate of this parameter. The regression also includes the parent quality and parent time variables from the model of Table 2. Finally, we also include variables measuring conventional school inputs: class size, the percentage of teachers in the school with over three years' teaching experience, and private school. Generally these variables performed poorly, explaining little and often being perversely signed. We report only the results for the class size variable, since this is usually identified as the premier candidate for social policy. We have also experimented with a number of different formulations of this variable (using a variety of dummies for small class or large class, and also interaction terms between class size and socioeconomic status or stream). In no cases were the results qualitatively different from those presented here.

The exact functional form of the relationship between attainment and the peer group is also an issue of substance in this work because, with diminishing returns to average peer quality, if policy is directed towards average attainment, then it is optimal policy to mix children together as much as possible. On the other hand, if there were increasing returns, it would be best to segregate them. We have discussed this elsewhere (Robertson and Symons 1996b). To investigate this issue, we experimented with including quadratic terms in this specification. The point estimates of the quadratic coefficients were always negative, indicating diminishing returns, significantly so for maths, but insignificantly for reading improvement. Fitting cubic terms, we found no evidence of regions of increasing returns. Henderson *et al.* (1978) have also found diminishing returns to the peer variable measured as average class IQ in a study of third grade children in the United States.

By and large, these results are not much different for maths and reading, and we concentrate on the maths results. For the unstreamed schools, the parental variables have a qualitatively similar effect on the improvement between 7 and 11 as they had on the attainment at age 7: the children of the educated and the successful do best. Family size is broadly as before. The absence of a father has a negative but insignificant effect. In contrast to early attainment, a working mother does not have an important effect on improvement between 7 and 11. Class size is incorrectly signed. The peer group variables, in contrast, are strong.

Turning to the streamed schools, note that both sets of peer group variables—membership of ability group and percentage of fathers in the top socioeconomic groupings—are important. Most other variables however have smaller parameters and reduced significance compared with unstreamed schools. We find that being in the top stream has a positive effect, and in the bottom stream a negative effect, on improvement between ages of 7 and 11 in both maths and reading.

There are some intriguing differences between streamed and unstreamed schools. In streamed schools scores in maths and reading at age 11 are significantly related to little other than the peer group variables, while for unstreamed schools parental quality remains extremely strong. One interpretation of this is that in streamed schools the values inculcated by a peer group comprised of children of similar abilities are much more potent than those derived from the family. Thus, early performance or early ability is much less important in a streamed school. One can interpret the differences between the parameters in Tables 3 and 4 as saying that the heightened peer group effects of the streamed school replace the flow of improved attainment derived from parental variables in non-streamed schools.

Endogeneity, measurement errors and instrumental variables

The results presented in Tables 3 and 4 and discussed above argue for the importance of peer group effects. If there are significant biases associated with the estimation, then these arguments are flawed. To see the various biases that might occur, consider the following schematic structure:

$$\Delta A = \alpha A + \beta s + \gamma p + \varepsilon \tag{3}$$

$$A^* = A + \upsilon \tag{4}$$

$$A_j^T = \theta_j A + \mu_j, \qquad j = 1, 2, 3 \ldots J \tag{5}$$

$$s = \sum_j \phi_j A_j^T + \lambda \tag{6}$$

$$p = \sum_k \delta_k r_k + \eta \tag{7}$$

where A is attainment at 7 and ΔA the improvement from 7 to 11. Equation (3) is then the educational production function and links the improvement to the child's initial attainment, stream s, and peer quality p. The term ε captures all unobserved influences. A is not directly observed, but is measured in the NCDS by test scores A^*. Equation (4) gives the relation between measured attainment and true A subject to the sampling error υ associated with any test score. Teachers assess child ability in J dimensions (in our case $J = 6$ as in Table 1) via (5), subject to errors μ_j. The allocation of children to streams is given by (6) depending on the teachers' assessment with error λ. Finally, (7) is a regression of peer quality on region of birth dummies r_k, wherein the error η is orthogonal to the r_k by construction.[3]

To deal with the problem of measurement error in (3), we use as instruments the A_j^T. This will be valid if $\mathrm{Cor}(\mu_j, \varepsilon) = 0$. Streaming is endogenous if $\mathrm{Cor}(\lambda, \varepsilon) \neq 0$, but once again the A_j^T will be valid instruments if $\mathrm{Cor}(\mu_j, \varepsilon) = 0$. Peer quality will be endogenous if $\mathrm{Cor}(\eta, \varepsilon) \neq 0$, but here regional dummies will be valid as instruments as long as $\mathrm{Cor}(r_k, \varepsilon) = 0$ and, as we have argued above, the r_k have explanatory power for p.

Consider first the plausibility of the assumption $\mathrm{Cor}(\mu_j, \varepsilon) = 0$. This will fail if, in forming their assessments of a student's ability in mathematics, say, teachers had some knowledge of the child's aptitude for higher work and allowed this to influence their assessment. This is certainly possible, and we seek to minimize the chance of this by excluding as instruments in the maths improvement regression the teachers' assessments associated with mathematical ability, and from the reading improvement regressions assessments of reading ability. This reasoning is confirmed by Sargan tests of instrument-residual orthogonality in that these fail if the exclusion is not made but are passed if the instrument set is so restricted.

The second orthogonality condition, $\mathrm{Cor}(r_k, \varepsilon) = 0$, fails if the mean of ε varies with region. How might this occur? First, there could be differences in attitudes to education between regions, either for inherent cultural reasons, or because like-minded parents congregate in certain regions to exploit good peer groups, for example. We regard the former as unlikely, given that we control for social class, parental educational levels, etc. Consider next the possibility of widespread regional choice by parents. The hypothesis is that ambitious parents will tend to move to areas where good schools are readily available for their children. The choice of broad region of birth for the instruments will itself mitigate this problem. But also, if there were such tendencies, there should be observable movement of families to regions of good schools following the birth of their first child. Of course, some families may move in anticipation of the birth of children, but there seems little reason to do this if they are attached to living in a particular region for other reasons: it would surely be unusual for people to move to areas of good schools several years before their as yet unborn child attains school starting age. In any case, even if some families did so move, it seems likely that large numbers would not. To investigate this in our data-set, for first-born children we fitted a logit model (not reported) giving the probability of an interregional move between birth and age 7 as a function of the interregional difference in the peer group variable and a measure of distance (0 for no move, 1 for a move to contiguous regions, 2 for moderate moves and 3 for moves across the country). The peer group effects were positive but insignificant. Thus, we conclude that there is at most only a slight tendency for parents to move to better regions *after* the birth of a first child, and hence that regional selection by ambitious parents is not an important phenomenon. These arguments indicate that, probably because of the coarseness of the regional measures, we can treat them as exogenous.

Second, $\mathrm{Cor}(r_k, \varepsilon) = 0$ fails if schooling provision varies regionally. However, many aspects of schooling (e.g. teachers' pay, syllabuses, length of school day) were determined centrally by government and national unions and will not have regional variation. Others, which might vary regionally for historical reasons (such as teachers' experience or class size), are observed in the NCDS and can be included. Essentially, given the other covariates, $\mathrm{Cor}(r_k, \varepsilon) = 0$ is not implausible in a culturally homogeneous unitary state such as England and Wales were in the 1960s: important differences may occur because of social class or parental education, but not because of location. Again, Sargan tests confirm our intuition.

Tables 5 and 6 report the instrumental variable regressions of the educational production function (equation (3) with the additional parental and schooling explanatory variables), treating the scores at age 7, the peer group effects and class size as endogenous variables, using regional dummies and teachers' assessments of child ability as instruments. The point estimates are broadly as before. Formal Hausman tests performed at this stage indicate no endogeneity problem and consequent validity of the OLS estimates of Tables 3 and 4. Such tests can have low power, and we show below that these data are characterized by important endogeneity (see footnote 5). The peer group effects remain correctly signed and marginally significant throughout, while the class size variable continues to perform badly. The point estimate of the coefficient on the peer group variable rises, suggesting the dominant problem in the estimates of Tables 3 and 4 arises from mismeasurement of variables rather than endogenous selection of peer groups (which would tend to move the parameter in the opposite direction). Instrumentation causes the parameters on initial ability to rise, consistent with measurement error biasing

TABLE 5

MATHS IMPROVEMENT, 7–11 TREATING PEER GROUP AND SCHOOL VARIABLES AS ENDOGENOUS

	Streamed school		Unstreamed school	
Maths score at 7*	−0.6	(3.7)	0.0	(1.3)
Peer group				
Top stream*	17.1	(1.7)	—	
Bottom stream*	−26.5	(2.5)	—	
% top SES*	0.21	(1.8)	0.15	(1.4)
Parent quality				
Father in top SES	−0.5	(0.2)	3.9	(1.7)
Father in middle SES	−0.3	(0.2)	2.8	(2.5)
Father stayed on	1.0	(0.6)	1.4	(1.0)
Mother stayed on	1.4	(0.9)	3.6	(2.9)
Parent time				
Mother works FT	0.7	(0.4)	−0.9	(0.7)
Mother works PT	2.3	(1.4)	0.5	(0.4)
No father	−2.3	(0.9)	−1.8	(0.7)
Family size	0.2	(0.3)	−1.6	(6.0)
School quality				
Class size*	0.11	(0.3)	0.40	(1.2)
Sex (=1 if girl)	−3.6	(2.3)	1.0	(1.2)
Scotland	0.8	(0.8)	8.3	(7.4)
Constant	12.3	(0.7)	−30.7	(2.6)
R^2	0.47		0.03	
No. of observations	1526		3322	

Notes: as for Table 3, except that variables marked with an asterisk are treated as endogenous. Instruments are teachers' assessments of student's abilities at 7 by the non-maths related criteria used in the regression reported in Table 1 and dummies for region of birth. *R*2 is calculated as the squared correlation between the actual and fitted dependent variable.

OLS estimates. Note that in general differences in attainment at age 7 are less persistent in the streamed schools.

The tables report the Sargan tests of the orthogonality of the instruments to equation error. These are passed at the 1% level in all four cases, and at the 5% level in three cases. The weakest results are for the maths improvement regressions in unstreamed schools. We consider this further below.

We must also verify that our instruments have explanatory power for the instrumented variables. To investigate this, we regressed the peer group variable on the set of instruments and predetermined variables for each of the four samples used in the regressions of Tables 3 and 4. *F*-statistics of tests of zero coefficients on the regional dummies gave values 6.45, 12.01, 6.70 and 12.04 for the four regressions (maths streamed, maths unstreamed, reading streamed and reading unstreamed, respectively). These are distributed as $F(9, 1456)$, $F(9, 3200)$, $F(9, 1457)$ and $F(9, 3223)$, respectively, under the null of no explanatory power for the instrumented variable. Bound *et al.* (1995) and Staiger and Stock (1997) suggest problems in instrumental variables regressions,

TABLE 6

READING IMPROVEMENT, 7–11 TREATING PEER GROUP AND SCHOOL VARIABLES AS ENDOGENOUS

	Streamed school		Unstreamed school	
Reading score at 7*	−0.5	(3.2)	−0.4	(22.8)
Peer group				
Top stream*	4.4	(0.6)	—	
Bottom stream*	−17.8	(1.6)	—	
% top SES*	0.22	(2.6)	0.12	(1.9)
Parent quality				
Father in top SES	−2.4	(1.1)	2.7	(2.1)
Father in middle SES	−1.4	(1.3)	0.9	(1.5)
Father stayed on	0.7	(0.6)	0.4	(0.4)
Mother stayed on	1.2	(1.0)	1.6	(2.3)
Parent time				
Mother works FT	−0.1	(0.1)	−0.1	(0.2)
Mother works PT	0.2	(0.2)	−0.3	(0.4)
No father	−1.0	(0.5)	−1.2	(0.9)
Family size	0.0	(0.1)	−0.7	(4.1)
School quality				
Class size*	0.43	(1.4)	0.23	(1.3)
Sex (=1 if girl)	−2.7	(3.5)	−3.5	(7.7)
Scotland	−1.7	(0.9)	−1.9	(3.1)
Constant	10.5	(0.6)	5.1	(0.8)
R^2	0.28		0.31	
No. of observations	1525		3318	

Notes: as for Table 5, except that instruments are teachers' assessments of student's abilities at 7 by the non-reading related criteria used in the regression reported in Table 1 and dummies for region of birth.

with such F-values as high as 5. In two of the four cases it may be that we have weak instrument problems. We shall investigate this further.

Some Monte Carlo investigation

Our regressions support the importance of the peer group effects in the educational production function. However, there is the possibility that our original regressions are misspecified with consequent biases, yet the diagnostic tests performed on the instrumented regressions have little power and the evidence from them should be treated with caution. This would often be the case in econometric investigations where the twin demands of instrument orthogonality and correlation with the endogenous variable can drastically reduce the available instruments. Moreover, reliance on asymptotic distributions for diagnostic tests may not provide reliable inference in finite samples, even for the large sample sizes that are now typical in applied work. In these circumstances, Monte Carlo investigation can resolve doubts about the validity of inference.

Write the model formally as

$$y = X\beta_1 + W\beta_2 + \epsilon$$

$$X = Z\pi + U$$

wherein y is a $T \times 1$ vector, X is a $T \times n$ vector of endogenous variables, W a $T \times k$ vector of predetermined variables, β_1 and β_2 are $n \times 1$ and $k \times 1$ coefficient vectors and ϵ a $T \times 1$ vector of errors whose distribution we discuss below. Z is a $T \times (r + k)$ set of instruments with $r > n$,[4] π an $(r + k) \times n$ matrix and $U = (u_1, u_2, \ldots, u_n)$ is $T \times n$. Define $G = [\epsilon, U]$. and let Ω be the $(n + 1) \times (n + 1)$ matrix with elements $\Omega_{ij} = E(G_{ti}G_{tj})$. Endogeneity is manifested as non-zero off-diagonal elements in the first row of Ω.

We select W and Z to match our sample data and β_1, β_2 to be the estimated coefficients of each regression of Tables 5 and 6 in turn. We construct the matrix π by performing the subsidiary regressions in our data. For the Monte Carlo, values of ϵ and $u_1, \ldots, u_n$ were drawn from a multivariate normal distribution with covariance matrix Ω, chosen to match the sample covariance matrix of the residuals from the reported regressions.[5] Using these generated errors, we construct values for X and y. Both OLS (of y on (X, W)) and instrumental variable (y on (X, W) using instruments Z) estimates of β_1 and β_2 are obtained. This Monte Carlo experiment was repeated 1000 times for each of the regressions reported in Tables 5 and 6.

Table 7 sets out the average bias and standard deviation of the endogenous coefficient estimates, the average calculated (nominal) standard errors from the repetitions of the IV estimates, the ratio of the calculated standard error of the IV estimates to the standard deviation and the root mean squared error of the OLS and IV estimates. It shows that the instrumental variables estimators are in general substantially less biased than the OLS estimates and always so for the endogenous variables. Bias in the IV estimates of the peer group effect is trivial in magnitude though sometimes statistically significant measured by the sample standard deviation of the Monte Carlo. Generally, OLS bias and IV bias are in the same direction, as noted by Bound *et al.* (1995). While on bias IV is clearly superior, its standard errors are worse by a factor of approximately 5, and in root mean squared error terms there is little to choose between the two estimation procedures as a general rule.

Turning to inference, note that the standard deviation of the Monte Carlo population of estimates is typically within two or three percentage points of the average estimated standard error computed from the usual IV formula. Thus, the reported asymptotic standard errors provide a good approximation to the small sample variation. We are also able to investigate the distribution of the estimates from the Monte Carlo. Thus, for the first Monte Carlo there are 1000 drawings for each of 16 estimated coefficients. Write the (1000×16) matrix of such estimates as B. Subtract the column mean from each column entry, and, under the hypothesis that the parameter estimates are multivariate normal, the 16,000 elements of the matrix BL (where L is the inverse of the Cholesky matrix of the variance–covariance matrix of B) are asymptotically independent $N(0, 1)$ variates. For each of the four Monte Carlo simulations, Kolmogorov–Smirnov tests accepted normality of the coefficients at a high P levels. Of course, from the

TABLE 7
MONTE CARLO RESULTS

	Truth	$bias_{OLS}$	$bias_{IV}$	$s.d._{OLS}$	$s.d._{IV}$	$s.e._{IV}$	ratio	$rmse_{OLS}$	$rmse_{IV}$
(a) Maths improvement in streamed school									
Score at 7	−0.57	−0.12	−0.01	0.02	0.14	0.14	0.97	0.12	0.14
Top stream	17.13	3.01	2.27	1.04	8.36	8.65	1.04	3.19	8.66
Bottom stream	−26.51	11.88	2.21	1.14	9.48	9.36	0.99	11.94	9.73
% top SES	0.21	−0.09	−0.01	0.02	0.11	0.11	1.02	0.09	0.11
Class size	0.11	−0.08	0.04	0.07	0.37	0.36	0.99	0.10	0.37
(b) Maths improvement in unstreamed school									
Score at 7	0.05	−0.56	−0.01	0.01	0.04	0.04	1.02	0.56	0.04
% top SES	0.15	−0.00	0.00	0.02	0.10	0.10	1.03	0.02	0.10
Class size	0.40	−0.32	−0.02	0.05	0.33	0.33	1.00	0.33	0.33
(c) Reading improvement in streamed school									
Score at 7	−0.55	−0.11	−0.02	0.02	0.15	0.15	1.02	0.11	0.15
Top stream	−4.43	12.60	3.36	0.80	6.89	7.20	1.04	12.63	7.67
Bottom stream	−17.83	9.82	2.80	0.86	9.70	9.66	1.00	9.86	10.10
% top SES	0.22	−0.12	−0.02	0.01	0.09	0.09	1.03	0.12	0.09
Class size	0.43	−0.39	−0.09	0.05	0.29	0.30	1.05	0.39	0.30
(d) Reading improvement in unstreamed school									
Score at 7	−0.40	−0.09	−0.00	0.01	0.02	0.02	0.99	0.09	0.02
% top SES	0.12	−0.07	−0.01	0.01	0.06	0.06	1.02	0.07	0.06
Class size	0.23	−0.18	−0.02	0.03	0.18	0.18	1.01	0.18	0.18

This table reports average bias in the endogenous variables of Tables 5 and 6 from 1000 Monte Carlo repetitions as described in the text. The columns report the true parameters generating the results, the average bias and standard deviation by OLS and IV, the average calculated (nominal) standard errors from the repetitions of the IV estimates, the ratio of the calculated standard error of the IV estimates to the standard deviation and the root mean squared error of the OLS and IV estimates.

perspective of inference, what matters is not so much the shape of the entire distribution as the weights in the tails. The empirical 0.5, 2.5, 5, 95, 97.5 and 99.5 percentiles from the four Monte Carlo simulations matched those of a normal variate to the first decimal place.[6]

On balance, these experiments indicate that the IV estimates are normally distributed, that the reported asymptotic standard error is a good guide to the true standard deviation, and that biases are generally small relative to that standard error. For the peer group variable, the direction of bias implies that the true size of the *t*-tests are a little lower than their nominal levels.

The above experiments set β_1 and β_2 to the observed IV estimates. We repeated the Monte Carlo, setting these parameters to the OLS estimates (reported in Tables 3 and 4). The main conclusions of bias and inference are unchanged, though OLS is here superior in terms of root mean squared error.

While the above analysis confirms that the instruments are adequately correlated with the endogenous variables, there is some evidence that, for the maths improvement in unstreamed schools regression, the instruments are not orthogonal to the equation error. (The *F*-statistic is significant at the 5% but not the 1% level.) We mimic this in Monte Carlo by running the subsidiary regression

$$e = Z\delta + v,$$

where e are the residuals from the IV, and replace the covariance matrix of (ϵ, U) with that of (v, U) to generate the random numbers. In each experiment the empirical version of ϵ is then generated as $Z\hat{\delta} + v$. The first term is held constant through the repetitions. This generates mild correlation between the instruments and the structural error of the same order as observed in the Sargan test. Monte Carlo indicates that this degree of correlation hardly changes the bias.

Educational attainment and earnings

We now turn to a set of forecasting regressions relating the attainment variables observed at age 11 to earnings at age 33. The dependent variable is the log of earnings per week for current or most recent job. We discard as outliers those earning less than £100 per week and those over £500 per week. We retain all right-hand-side variables from the improvement regressions to capture any on-going effects from these variables. Table 8 reports results separately for males and females. Note first that the effect of educational attainment is similar between males and females. Both reading and maths at 7 and the improvements by age 11 are richly rewarded. For males, there is some

TABLE 8
EARNINGS REGRESSIONS

	Males		Females	
Maths score at 7	0.0026	(4.1)	0.0040	(4.4)
Maths improvement	0.0011	(2.0)	0.0028	(3.4)
Reading score at 7	0.0039	(4.2)	0.0032	(2.4)
Reading improvement	0.0015	(2.0)	0.0026	(2.1)
% top SES	0.0006	(1.2)	−0.0003	(0.5)
Father in top SES	0.0812	(2.4)	−0.0015	(0.0)
Father in middle SES	0.0911	(4.1)	0.0039	(0.1)
Father stayed on	0.0245	(1.0)	0.0680	(2.0)
Mother stayed on	0.0501	(2.1)	0.0158	(0.5)
Mother works FT	−0.0377	(1.3)	−0.0156	(0.4)
Mother works PT	0.0000	(0.0)	−0.0310	(0.7)
No father	0.0027	(0.0)	0.0688	(0.7)
Family size	−0.0104	(1.6)	−0.0171	(1.7)
Class size	0.0010	(0.8)	−0.0028	(1.5)
Constant	5.2031	(74.0)	5.0761	(49.8)
R^2	0.18		0.14	
No. of observations	1130		783	

Notes: as for Table 3. Dependent variable is ln(weekly earnings at age 33).

evidence of ongoing effects beyond age 11 of parental SES variables, with little measured effect for females. This differential effect is something of a mystery. There is some evidence that number of siblings has a negative effect on earnings. The evidence for class size is mixed: for males this variable is wrong-signed and insignificant; for females it is correctly signed and marginally significant. We experimented also here with instrumenting the class size variable: again this result was qualitatively unchanged. Overall, we find little evidence for the widely held belief that smaller classes offer a better environment for educational achievement.

Attrition in the NCDS

Of the original 18,359 individuals eligible for inclusion in the NCDS, only some 12,000 remain at age 33. Additionally, the full set of variables is not uniformly available even for this subsample, so that for example our earnings regressions are estimated on only about 2000 individuals. Such severe attrition must give cause for concern. One could in principle model the probability of response and the regression models jointly, conditioned on exogenous variables. However in the NCDS this is not possible, as there is no set of exogenous variables that exist for all individuals in the original cohort. There thus seems no way formally to correct the regression results for possible non-random attrition bias. The extent of the problem can be gauged from Table 9, which summarizes the distribution of maths scores at age 7 for those individuals included in regressions at ages 7, 11 and 33. It is immediately apparent that we are undersampling the lower tail of the distribution as the cohort ages by this measure. At age 11, when our score improvement regressions are performed, this seems to be a relatively minor problem. At age 33, where the increase in mean is of greater significance, part of this increase is of course due to the fact that those employed tend to be drawn from the upper portion of the maths score distribution. In fact, the mean maths score at age 7 of those included in NCDS but not reporting earnings at age 33 is about two points below the mean at age 7 recorded in Table 9. Thus, the high mean for those included in the

TABLE 9

ATTRITION DISTRIBUTIONS

Distributions of maths score at age 7 for regression samples at ages 11 and 33

	Score							
Age	0–20	21–40	41–60	61–80	81–100	Mean	s.d.	No. obs.
7	14.58	25.97	27.60	21.31	10.55	52.35	24.27	8733
11	14.73	25.43	27.37	22.01	10.44	52.48	24.41	4848
33	13.70	24.83	26.87	22.79	11.81	53.84	24.50	1913

Note: The first row gives the percentage distribution, the mean and the standard deviation of maths scores at age 7 for those individuals included in the regression of Table 2, the second row gives the same for those individuals included in the regressions of Table 3 and the third row for those individuals included in the regressions of Table 6.

regressions at age 33 is likely to be due in large part to this effect rather than to genuine attrition.

III. Peer Group Effects and Streaming

There are two peer group effects studied in this paper: the effect of classmates coming from higher socioeconomic groups, and the effect caused within schools by streaming. Students benefit from being placed in the top stream of streamed schools, but suffer if they are placed in lower streams. Figure 6 graphs the expected maths score at age 11 as a function of score at age 7 for students in unstreamed schools and in the three possible streams of streamed schools. Those placed in the top stream benefit from attending streamed schools at all initial maths scores, with the greatest benefit seen at lower scores—clearly supportive of a peer group effect. Attendance at a streamed school and being placed in the lower-ability stream provides a better outcome than attending an unstreamed school only for those with very low initial scores. Again, this supports the peer group effect, the average quality of peers in an unstreamed school presumably exceeding that of the lowest stream in a streamed school.

Apart, therefore, from a small group of low-ability maths students, the general pattern from both instrumented and uninstrumented regressions for both maths and reading is that streaming improves the outcomes for the strong students and worsens outcomes for weak students. This conclusion is in line with conventional wisdom and in contrast to the recent findings of Figlio and Page (1999). Whether or not schools should stream by ability thus involves trading off the gains to the able against the losses to the less able. As discussed

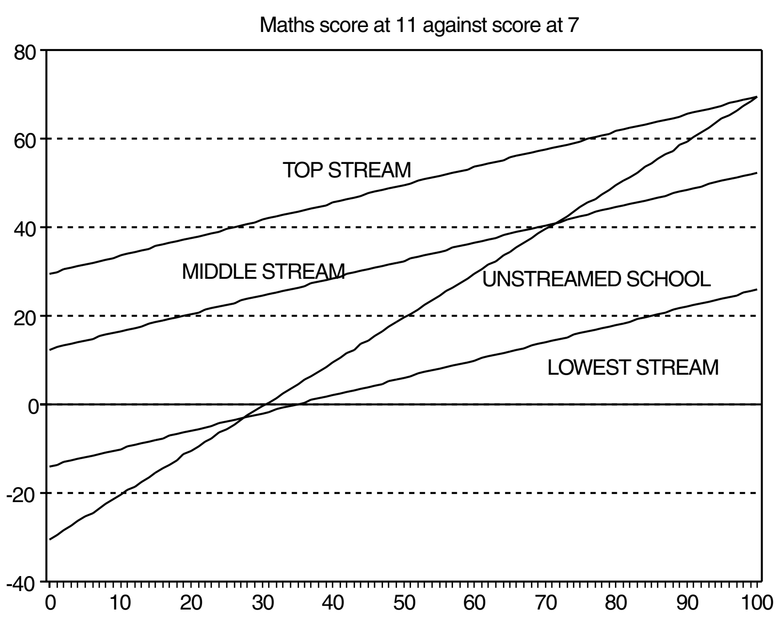

Figure 6

above, diminishing returns to the SES peer group would imply mixing for utilitarian policy objectives. It is not clear how one might achieve such mixing. Forms of coercion such as busing are widely discredited and may antagonize those with higher ability. A free market in schooling wherein fees reflect marginal contributions to the joint production of human capital may attain the optimum (for a demonstration see Robertson and Symons 1996b), but such a system could involve severe equity issues, with poor students having to pay higher fees.

IV. CONCLUSIONS

This paper has studied the determinants of academic attainment at age 11 and the links between this attainment and earnings at age 33. We have found strong evidence for the importance of peer groups. We have found also strong effects from parental social class and parental academic achievement. While we are not able to trace the exact channel by which the peer group influences attainment, it seems likely to us that a better peer group brings with it better behaved children, a belief in the value of education, and parents who actively scrutinize the teaching process.

To see the strength of these effects, it is instructive to calculate the implied increment in expected earnings at age 33 for a male with father in the top socio-economic group, whose parents both stayed at school beyond minimum leaving age, and who attended an unstreamed school in which half of their fellow pupils have parents in the top social class. Using our estimated models, we find that reading and maths scores at age 7 rise by 19.4 and 17.0, respectively. At age 11 this, plus ongoing effects from parents and peers, raises scores by 22.3 and 33.4 points in reading and maths, respectively. The earnings regressions translate these increases into an extra 33% in income relative to an individual possessing the lowest possible values of these attributes.

ACKNOWLEDGMENTS

This is a substantial rewriting of Robertson and Symons (1996a). Both authors are at the Centre for Economic Performance, LSE. We are grateful to the ESRC Data Archive for providing access to the NCDS data, to Leon Feinstein, Hashem Pesaran and Nicholas Rau, and to two anonymous referees for extremely helpful suggestions, and to seminar participants at UCL, LSE and Oxford University.

NOTES

1. Other parameter estimates were almost always within one standard error of each other.
2. Note however that our specification does not suffer from the problem of identification of peer group effects discussed in Manski (1993), wherein averages of the dependent variable are included as explanators.
3. There are 11 regions: North-West, Northern, East and West Yorkshire, North Midlands, Eastern, Midlands, London and the South East, Southern, South West, Wales and Scotland. Scotland is included as explanatory variable, and one region is dropped to prevent multicollinearity with the constant of the regression.
4. In the just identified case, Monte Carlo may not be informative, as the moments of the instrumental variable estimator will not exist.

5. The implied correlation matrix was generally found to have non-trivial off-diagonal elements in the first row, indicating endogeneity of the X variables. The values were:

math str	−00.28	−0.12	−0.05	−0.14	0.36
math unstr	−0.65	−0.03	−0.08		
read str	−0.14	−0.19	0.18	0.23	0.19
read unstr	−0.19	−0.11	−0.10		

The order of the equations for the endogenous variables in the streamed regressions is score at 7, % top SES, class size, top stream and bottom stream. For the unstreamed regressions the last two are absent. Standard errors for these Spearman correlation coefficients are of order 0.02, indicating substantial endogeneity, contrary to the conclusions of the Hausman tests above.

6. The values were:

	0.5%	2.5%	5%	95%	97.5%	99.5%
Monte Carlo 1	−2.68	−1.94	−1.64	1.62	1.97	2.62
Monte Carlo 2	−2.59	−1.97	−1.63	1.62	1.98	2.59
Monte Carlo 3	−2.67	−1.96	−1.63	1.62	1.96	2.64
Monte Carlo 4	−2.58	−1.95	−1.65	1.65	1.95	2.56
$N(0, 1)$	−2.58	−1.96	−1.65	1.65	1.96	2.58

REFERENCES

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.

BECKER, G. S. (1981). *A Treatise on the Family*. Cambridge, Mass.: Harvard University Press.

BETTS, J. R. (1995). Does school quality matter? Evidence from the National Longitudinal Survey of Youth. *Review of Economics and Statistics*, **77**, 231–50.

BORJAS, G. J. (1992). Ethnic capital and intergenerational mobility. *Quarterly Journal of Economics*, **107**, 123–50.

BOUND, J., JAEGER, D. A. and BAKER, R. (1995). Problems with instrumental variable estimators when the correlation between the instruments and the endogenous explanatory variable is weak. *Journal of the American Statistical Association*, **90**, 443–50.

CARD, D. and KRUEGER, A. (1992). Does school quality matter? *Journal of Political Economy*, **100**, 1–40.

COLEMAN, J. S. *et al.* (1966). *Equality of Educational Opportunity*. Washington DC: US GPO.

EPPLE, D. NEWLON, E. and ROMANO, R. (2000). Ability tracking, school competition and the distribution of educational benefits. NBER Working Paper No. 7854.

EVANS, W. N., OATES, W. E. and SCHWAB, R. M. (1992). Measuring peer group effects: a study of teenage behaviour. *Journal of Political Economy*, **100**, 966–91.

FEINSTEIN, L. ROBERTSON, D. and SYMONS, J. (1999). Nursery education and attainment in the NCDS. *Education Economics*, **7**, 209–34.

FIGLIO, D. N. and PAGE, M. E. (1999). School choice and the distributional effects of ability tracking: does separation increase equality? NBER Working Paper No. 8055.

HANUSHEK, E. A. (1986). The economics of schooling: production and efficiency in public schools. *Journal of Economic Literature*, **24**, 1141–77.

—— (1992). The trade off between child quantity and quality. *Journal of Political Economy*, **100**, 84–117.

HARBISON, R. W. and HANUSHEK, E. A. (1992). *Educational Performance of the Poor: Lessons from Rural North-east Brazil*. Oxford University Press for the World Bank.

HENDERSON, V. MIESZKOWSKI, P. and SAUVAGEAU, Y. (1978). *Peer Group Effects and Educational Production Functions*. Ottawa: Economic Council of Canada.

HOXBY, C. (2000). Peer effects in the classroom. NBER Working Paper No. 7867.

MANSKI, C. (1993). Identification of endogenous social effects: the reflection problem. *Review of Economic Studies*, **60**, 531–42.

ROBERTSON, D. and SYMONS, J. (1996a). Do peer groups matter? Centre for Economic Performance LSE Discussion Paper No. 311.

—— and —— (1996b). Self-selection in the state school system. Centre for Economic Performance LSE Discussion Paper No. 312.

Sacerdote, B. (2000). Peer effects with random assignment. NBER Working Paper No. 7469.
Staiger, D. and Stock, J. (1997). Instrumental variable regression with weak instruments. *Econometrica*, **65**, 557–86.
Summers, A. A. and Wolfe, B. L. (1977). 'Do schools make a difference? *American Economic Review*, **67**, 639–52.