Is Free School Meal Status a Valid Proxy for Socio-Economic Status (in Schools Research)?

Graham Hobbs

Anna Vignoles

May 2007

Published by
Centre for the Economics of Education
London School of Economics
Houghton Street
London WC2A 2AE

May 2007
ISBN 978-0-85328-197-9

The Centre for the Economics of Education is an independent research centre funded by the Department for Children, Schools & Families. The views expressed in this work are those of the author and do not reflect the views of the DCSF. All errors and omissions remain the authors.

Executive Summary

Across the social sciences, family socio-economic status (SES) is seen as a potentially key determinant of children's educational attainment. However, a lot educational research in the UK relies on administrative datasets that rarely contain measures of SES. Instead, they almost always include an indicator of pupils' "Free School Meal (FSM) Eligibility". FSM status is widely used as a proxy for SES in UK educational research, therefore. Future research is likely to rely even more heavily on administrative data, in particular, the National Pupil Database (NPD), and hence on FSM status as a proxy for SES. There is therefore a pressing need to evaluate the validity of the FSM measure in this context. This is our aim.

The "FSM eligibility" measure is, in fact, a measure of claiming FSM, rather than just eligibility. Nevertheless, our evaluation begins by examining FSM eligibility rules and nationally-representative data on those eligible for FSM. Children in families claiming Income Support (IS) and Income-Based Job Seekers Allowance (IB-JSA) were eligible for FSM in 2001/2 (the year relevant to our main analysis). Broadly to be eligible for these benefits in that year, a child must have been in a household without a member working more than 24 hours a week, with a low income (defined relative to needs) and limited capital assets.

National data suggests that the majority of children claiming free school meals will be dependents of IS claimants (rather than IB-JSA claimants). National data also suggests the majority will be in "one parent families". Specifically, in May 2002, over 75% of primary school age dependents of IS claimants were in "one-parent families". National data also suggests that a significant proportion of children who are eligible to receive free school meals will remain eligible for more than five years. For example, 34% of "lone parent" claimants received IS for five or more years. Furthermore, administrative data suggests that only 10% of the cohort of children entering reception year in 1997/98 in state schools in England changed FSM status during the period of interest for this study (between January 2002 and January 2004).

The second part of our evaluation examines the relationship between FSM status *and* other key characteristics of the child's family. We do this by using a data set that includes both FSM status and rich data on the child's family background, namely merged Avon

Longitudinal Study of Parents and Children (ALSPAC) - National Pupil Database (NPD) data. These merged Avon data include not only FSM status of the child but also data on his or her family income, mother's and partner's employment, family employment, one-parent family status, mother's and partner's education, and mother's and partner's social class. The Avon data do paint a somewhat different picture to that described above from the national data. In particular, in our Avon data, 43% of children claiming FSM are in families with at least one parent in full-time employment and only 28% are in one-parent families. One explanation of these differences is that some families misreport their employment and partnership status when claiming benefits. Another explanation highlights a limitation of the data used here: the Avon measures of socio-economic status (SES) are observed five or more years before the FSM measure (from NPD) and so may genuinely change over time.

With the above caveat in mind, we investigated the range of SES measures to determine which measures FSM status proxies "best". We find that FSM status "best" proxies children in households with family incomes below £200 per week, i.e. the bottom quartile of the income distribution, rather than the bottom decile (even though less than 10% of children claim FSM in this sample). However, we do observe apparent anomalies. Firstly, 22% of FSM children have incomes above £200 per week and secondly, 20% of non-FSM children have incomes below £200 per week. Thus FSM status does not appear to identify all low-income children in our data, although we acknowledge the problem above about the timing of the data we have.

For family employment, FSM status proxies "best" children in workless families and those with only one part-time worker. While only 8% of non-FSM children are in workless families and those with one part-time worker, 43% of FSM children are in families with one or more full-time workers, or two part-time workers.

FSM status is a far from perfect proxy of one parenthood. In particular, 72% of FSM children have two-parents. FSM status is also a far from perfect proxy of mothers' and partners' education and social class.

We then investigated the use of FSM as an imperfect proxy for true SES in two specific contexts. Firstly, when researchers are trying to investigate differences in educational achievement for low and high SES children, where FSM status is the variable of interest.

Secondly, when researchers want to simply take account of family background and SES in their model and therefore include FSM status as a control variable in an OLS regression. We find that the bias produced by using FSM instead of true SES is context-specific.

Firstly, we assessed the extent of imperfect proxy bias when estimating differences in Key Stage 2 attainment by family income, using FSM status to proxy incomes below £200 per week. FSM status is the variable of interest, here. In English, maths and science, the difference in the average Key Stage 2 attainment of children with family incomes below and above £200 per week is 0.5 standard deviation units. However, the difference in the average Key Stage 2 attainment of FSM and non-FSM children is 0.6-0.7 standard deviation units. The difference in these differences is the imperfect proxy bias. The bias is around 30-40% of the "true" difference and is significant at the 1% level.

Finally, we assess the bias in a model of the effects of school type and Special Educational Needs status on a child's Key Stage 2 attainment. FSM status is a control variable, here. We estimate three specifications: 1) a "true" regression including our SES measures, 2) an omitted variables regression just omitting the SES measures, and 3) a proxy variable regression omitting the SES measures but including FSM status. The omitted variables regression equation suffers from omitted variables bias because it does not allow for the effects of SES. The proxy variable regression suffers from imperfect proxy bias because it includes only a proxy rather than true measures of SES. The amount of this bias is the difference in the parameter in the proxy variable and "true" regressions.

FSM is status is statistically significant in the proxy variable regressions, so it does predict achievement in the absence of true SES measures. However, FSM status does not always do a good job as a proxy for the true SES of the child. In general using FSM status reduces the bias caused by omitting SES altogether by only 10-25%. Furthermore, using FSM status instead of true SES generates imperfect proxy bias, which in our model was substantial on the voluntary-aided and voluntary-controlled schools coefficients. The imperfect proxy bias on these coefficients is 90% and 75-80% of the omitted variables bias, respectively (the reference category is community schools). On Special Educational Needs, it is 75-80%. So the biases arising from using FSM status instead of true SES are significant and sometimes large. Moreover, using FSM status instead of true SES sometimes changes the key findings from the model. For instance, the statistical significance of the school type effects are

sometimes different in the "true" and proxy variable regressions, influencing the basic interpretation of the results. For example, in maths and science, the effect of voluntary-aided schools is positive and significant in the proxy variable regression, but *insignificant* in the "true" regression.

Future research should evaluate the FSM measure in other contexts and datasets, and the use of various small area data matched to children's home postcodes as proxies for measures of SES. In the meantime, researchers should be cautious in drawing inferences from research reliant on the FSM measure. When used as the variable of interest, FSM status is an imperfect proxy of low income or "workless" families, or one-parenthood. In the context of estimating differences in educational attainment by family income, the bias generated by using FSM as opposed to "true" SES is quite large. When used as a control variable in an OLS regression, FSM status reduces omitted variables bias to a moderate extent only. In other words, if omitted variables bias is a concern, then the inclusion of FSM status in the model should do little to diminish this concern.

Is Free School Meal Status a Valid Proxy for Socio-Economic Status (in Schools Research)?

Graham Hobbs

Anna Vignoles

Acknowledgments

The authors would like to acknowledge the helpful and constructive comments of those who attended the Bedford Group seminar and the PLUG seminar in Bristol, at which this paper was presented. Particular thanks go to Rebecca Allen and Nikos Tzavidis for their input into this paper.

Graham Hobbs is a Research Officer for the Centre for the Economics of Education at the Bedford Group for Lifecourse and Statistical Studies, Institute of Education. Anna Vignoles is a Reader in Economics of Education at the Bedford Group for Lifecourse and Statistical Studies, Institute of Education, a Research Associate at the Centre for Economic Performance, London School of Economics and Deputy Director of the Centre for the Economics of Education.

1. Introduction

Across the social sciences, family socio-economic status (SES) is seen as a potentially key determinant of children's educational attainment.[5] However, a lot educational research in the UK relies on administrative datasets which rarely contain measures of SES. Instead, they almost always include an indicator of pupils' "Free School Meal (FSM) Eligibility". FSM status is widely used as a proxy for SES in UK educational research, therefore. Future research is likely to rely even more heavily on administrative data, in particular, the National Pupil Database (NPD), and hence on FSM status as a proxy for SES. There is therefore a pressing need to evaluate the validity of the FSM measure in this context. This is our aim.[6]

In some research, FSM status is the "variable of interest". In this context, it is typically used to proxy low income. One example is research on differences in educational attainment and progress by FSM status (e.g., Sammons et al., 1997a; Strand 1999; DfES 2003; DfES 2005b). Another example is research on the effects of school composition on educational attainment. Here the proportion of FSM pupils proxies the proportion of low-income pupils at school-level (e.g., Sammons et al., 1997b; Strand 1997; DfES 2003; Hutchison 2003; Schagen and Schagen 2005). A third example is research on socially-segregated schooling. Here, FSM status proxies low income in measures of segregation (e.g., Gibson and Asthana 2000; Noden 2000; Gorard et al., 2002, 2003; Goldstein and Noden 2003; Allen and Vignoles 2006).

In other research, FSM status is a "control variable". In particular, FSM status is included in models of educational attainment to eliminate or reduce the extent of omitted variables bias. In this context, FSM status is used, sometimes implicitly, as a proxy not just for income, and not just low income, but also for other unobserved SES variables. Important examples have included:

[5] Appendix 1 provides a brief review of evidence on the effects of SES variables on educational attainment in the UK.

[6] There has been little formal evaluation of the FSM measure. Croxford (2000) is a notable exception.

1. Studies of "ethnicity gaps" in educational attainment (Strand 1999; DfES 2003);
2. Studies of the effects of pupil mobility (Strand 2002);
3. Evaluations of education policies, e.g., Excellence in Cities (Machin et al., 2004) and the Literacy Hour (Machin and McNally 2004);[7]
4. Studies of selective versus comprehensive school systems (Atkinson et al., 2006);
5. Studies of the effects of class size (Blatchford et al., 2003);
6. Studies of the effects of school resources (Levacic et al., 2005); and
7. School effectiveness research (Strand 1997; Thomas et al., 1997).

The first part of our evaluation examines FSM eligibility rules and nationally-representative data on those eligible for FSM. The second part examines the joint distributions of FSM status and nine SES measures: family income, mother's and partner's employment, family employment, one-parent family status, mother's and partner's education, and mother's and partner's social class. We report the binary indicator of each SES measure which FSM status proxies "best", along with the associated probabilities of "false positives" and "false negatives".

The extent of imperfect proxy bias is context-specific. The third part of our evaluation examines two contexts, one when FSM status is the variable of interest, the other when FSM status is a control variable in an OLS regression. First, we assess the extent of bias when estimating differences in Key Stage 2 attainment by family income, using FSM status to proxy incomes below £200 per week. FSM status is the variable of interest, here. Second, we assess the extent of bias when estimating the effects of school type and Special Educational Needs on Key Stage 2 attainment, using FSM status to proxy eight SES measures. FSM status is a control variable, here.

Section 2 presents the conceptual framework. Section 3 discusses past evaluations. Sections 4 and 5 describe the method and data. Section 6 examines FSM eligibility rules and nationally-representative data on those eligible for FSM. Section 7 presents the main results. Section 8 offers concluding remarks.

[7] The control variable in these papers is the percentage of FSM pupils at school-level.

2. Conceptual Framework

Suppose a variable z is a proxy for a variable q. This section addresses two questions: Under what conditions is z a perfect proxy for q when z is the "variable of interest"? Under what conditions is z a perfect proxy for q when z is a "control variable" in an OLS regression?[8]

The proxy is the "variable of interest"

Suppose the proxy is binary, as is the case with FSM status, and is the variable of interest. A binary proxy cannot be a perfect measure of a continuous or categorical variable. It can be a perfect measure of a binary indicator of a continuous or categorical variable, however. For example, FSM status would be a perfect measure of family incomes below £100 per week if all FSM children have incomes below £100 per week and all non-FSM children have incomes above £100 per week. The imperfectness of FSM status as a proxy for this binary income variable is defined by two errors, the probability that FSM children have incomes above £100 per week ("false positives"), and the probability that non-FSM children have incomes below £100 per week ("false negatives").

More formally, z is a perfect proxy for a binary indicator of a continuous or categorical variable q if the distribution of q conditional on z = 0 and the distribution of q conditional on z = 1 do not overlap. In other words, z is a perfect proxy for a binary indicator of a continuous variable q if $q > q^*$ when z = 0 and $q \leq q^*$ when z = 1, for some constant q^*, and z is a perfect proxy for a binary indicator of a categorical variable q if q takes on the values $\{q^*+1, q^*+2, \ldots, J\}$ when z = 0 and q takes on the values $\{1, 2, \ldots, q^*\}$ when z = 1, for positive integers $q^*<J$.[9] If the distribution of q conditional on z = 0 and the distribution of q conditional on z = 1 overlap, then z is an imperfect binary proxy for every binary indicator of q.[10] The imperfectness of z as a proxy for a specific binary indicator of q, q_b=0 if $q > q^*$ and q_b=1 if $q \leq q^*$, is defined by the probabilities of false positives and negatives, Prob($q > q^*$| z

[8] The answer to the second question is based on Wooldridge (2002, p63-67).

[9] z can only be a perfect binary proxy for q if Prob($q \leq q^*$) equals Prob(z = 1).

[10] Obviously, if the distribution of q when z = 0 and the distribution of q when z = 1 are identical, then z is uninformative about q.

= 1) and Prob($q \leq q^* | z = 0$), respectively. Finally, the binary indicator of q which z proxies "best" is defined as that which minimises the sum of these probabilities of false positives and negatives.[11]

The proxy is a "control variable" in an OLS regression

Consider, for example, estimating the effects of school type on educational attainment in an OLS regression, omitting SES variables. If the SES variables are partially correlated with educational attainment, as the literature suggests, and if school type is partially correlated with the SES variables, then the parameters on school type are biased. This is omitted variables bias.

More formally, suppose that the "true" model of an outcome variable y is:

$$E(y | x_1,\ldots, x_K, q) = \beta_0 + \beta_1 x_1 + \ldots + \beta_K x_K + \gamma q$$

where $\mathbf{x} = (x_1, \ldots, x_K)$ are observed explanatory variables and q is an unobserved variable. Writing this in error form:

$$y = \beta_0 + \beta_1 x_1 + \ldots + \beta_K x_K + \gamma q + v \qquad (1)$$

Consider the consequences of ignoring the unobserved variable q. Write the linear projection of q on 1, **x** as:

$$q = \delta_0 + \delta_1 x_1 + \ldots + \delta_K x_K + r$$

Then the OLS regression of y on 1, **x** yields:

$$y = (\beta_0 + \gamma\delta_0) + (\beta_1 + \gamma\delta_1)x_1 + \ldots + (\beta_K + \gamma\delta_K)x_K + (v + \gamma r) \qquad (2)$$

[11] This is the value of q* which minimises [Prob($q > q^* | z = 0$) + Prob($q \leq q^* | z = 1$)].

The "omitted variable bias" on x_k is $\gamma\delta_k$. This result is well-known.

Returning to the example, suppose we have a proxy, FSM status, for the omitted SES variables. Under what conditions are the parameters on school type unbiased if FSM status is included in the regression? There are two conditions. The first is that the coefficient on FSM status is zero in the "true" regression including the SES variables.[12] The second is that school type is partially uncorrelated with the SES variables once we include FSM status in the regression. Both conditions are model-specific.

More formally, suppose that z is a proxy for q. Under what conditions does the OLS regression of y on 1, **x** and z produce consistent estimates of $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_K)$? The first condition is that the proxy variable is redundant in the "true" model. This can be stated as follows:

$$E(y \mid \mathbf{x}, q, z) = E(y \mid \mathbf{x}, q)$$

In other words, z is irrelevant for explaining y, in a conditional mean sense, once **x** and q have been controlled for. This assumption is virtually always made and is rarely controversial.

The second condition is that the correlation between the unobserved variable q and each x_k be zero once we partial out z. This is easily stated in terms of a linear projection:

$$L(q \mid 1, x_1, \ldots, x_K, z) = L(q \mid 1, z)$$

Write the linear projection of q on 1, **x**, z as:

$$q = \theta_0 + \rho_1 x_1 + \ldots + \rho_K x_K + \theta_1 z + r$$

Then the OLS regression on y on 1, **x** and z gives:

$$y = (\beta_0 + \gamma\theta_0) + (\beta_1 + \gamma\rho_1)x_1 + \ldots + (\beta_K + \gamma\rho_K)x_K + \gamma\theta_1 z + (\gamma r + v) \qquad (3)$$

[12] In other words, the SES variables, but not FSM status, determine educational attainment.

If both conditions hold, then the OLS regression of y on 1, **x** and z produces consistent estimates of **β**. The first condition means that z is uncorrelated with v and, by definition, z is uncorrelated with r. The second condition means that $\boldsymbol{\rho} = (\rho_1, \ldots, \rho_K) = \mathbf{0}$, and, by definition, **x** is uncorrelated with r. If the second condition fails to hold, then OLS is inconsistent. The "imperfect proxy bias" on x_k is $\gamma\rho_k$. The hope is that ρ_k is smaller in magnitude than if z were omitted from the linear projection (in other words, $|\rho_k| < |\delta_k|$).[13]

3. Past Research

There has been much criticism, but few evaluations, of the FSM measure. Shuttleworth (1995) assessed FSM status as a proxy for "social deprivation" in Northern Ireland. Non-Catholic children were less likely to be eligible for FSM (22%) than Catholic children (44%). In addition, children with both parents employed were less likely to be eligible for FSM (13%), than those with only the father employed (25%), or those with only the mother employed (31%), or those with both parents unemployed/inactive (67%).[14] In terms of the conceptual framework, FSM status proxies being Catholic, but the probabilities of false positives (38%) and negatives (37%) are high.[15] It also proxies having both parents unemployed/inactive, but while the probability of false negatives is low (11%), the probability of false positives is high (47%).

The focus of the paper was on the validity of FSM status when estimating the effects of social deprivation on GCSE attainment. FSM status was the variable of interest, therefore. In a multi-level model of GCSE attainment, *both* FSM status and family employment were

[13] This framework extends straightforwardly to the case of multiple unobserved variables. This extension permits the treatment of unobserved categorical variables as multiple unobserved binary variables. See appendix 2. This extension is not based on Wooldridge (2002).

[14] Children with no siblings were also less likely to be eligible for FSM (14%), than those with 1-2 siblings (22%), or those with 3-4 siblings (40%), or those with 5 or more siblings (53%).

[15] Own calculations from Shuttleworth (1995, table II). In other words, the probability of being Non-Catholic conditional on being eligible for FSM is 38%, and the probability of being Catholic conditional on not being eligible for FSM is 37%.

statistically significant, but being Catholic and family size were not. Shuttleworth (1995, p499-503) concluded, therefore, that FSM status was a "useful", but not, on its own, the "optimal" measure of SES (unsurprisingly).

Croxford (2000) assessed *school-level* free meal entitlement (FME) as "a valid measure of school intake characteristics" in Scottish schools.[16] The first part of the paper examined the validity of school-level FME as a measure of *school-level* SES. In an OLS regression, two-thirds of the variation in school-level FME was explained by five school-level SES measures.[17] Croxford concluded that the remaining one-third of the variation was explained by non-SES factors and, therefore, that school-level FME was an "inconsistent" measure of school-level SES (p333). More generally, Croxford argued that no single measure could account for all the variation in school-level SES.

The second part of the paper assessed school-level FME as a proxy for *pupil- and neighbourhood-level* SES in the context of estimating the between-school variance in Standard Grade attainment at the end of compulsory education. School-level FME was a control variable, therefore.[18] The paper reported the estimated between-school variance in both a multi-level regression of attainment on school-level FME, and a multi-level regression of attainment on pupil- and neighbourhood-level SES measures and school-level FME.[19] For one attainment measure, the imperfect proxy bias was approximately 10% (own calculations from Croxford 2000, table 5).[20] Croxford concluded that FME was a "relatively good

[16] The FME measure in Scotland is equivalent to "FSM Eligibility" in England and Wales. Unfortunately, the study had no data on pupil-level FME.

[17] More precisely, the dependent variable was ln((p/1–p)), where p was the (school-level) proportion of pupils entitled to free meals, and ln denotes the natural logarithm. The five school-level SES variables were constructed from pupil-level data on parental education, social class and employment, and home ownership, family size and one-parent family status, and (school) local area data on unemployment, population density and deprivation, by principal components analysis.

[18] The between-school variance is a measure of the size of school effects.

[19] The pupil- and neighbourhood-level SES variables were home ownership, family size, mother's and father's, education, social class and employment (all pupil-level), and the Carstairs deprivation score (neighbourhood-level).

[20] This is the difference in the estimated between-school variances in the two models as a percentage of the estimate in the second model. This is approximately the imperfect proxy bias because school-level FME is insignificant in the second model. This suggests that the first condition on the proxy variable is fulfilled.

surrogate" in this model (p331). For another attainment measure, the imperfect proxy bias was approximately 27% (own calculations), but the between school variance was insignificant in both models. Croxford concluded that FME was "not an adequate measure" in this model (p331).

4. Method

There are three parts to the analysis. The first part reports the joint distributions of FSM status and nine measures of SES: family income, family employment, mother's and partner's employment, one-parent family status, mother's and partner's education, and mother's and partner's social class. We report the binary indicator of each SES measure which FSM status proxies "best" (as defined in section 2.1), along with the associated probabilities of false positives and negatives. For example, does FSM status proxy "best" family incomes below £100 or £200 per week? If FSM status proxies "best" family incomes below £200 per week, what proportion of FSM children have incomes above £200 per week (false positives) and what proportion of non-FSM children have incomes below £200 per week (false negatives)?

The extent of "imperfect proxy bias" when using FSM status as the variable of interest is context-specific. The second part of the analysis assesses the extent of this bias when estimating differences in Key Stage 2 (KS2) attainment by family income. The bias is estimated when FSM status is used to proxy family incomes below £200 per week. This is the binary indicator of family income which FSM status proxies "best" in our data. In particular, we estimate:

(1) The difference in average KS2 attainment of children with incomes below and above £200 per week; and
(2) The difference in average KS2 attainment of FSM and non-FSM children.

The imperfect proxy bias is the difference in these differences. Bootstrapping is used to estimate the standard error of this bias and test if it is statistically significant.[21]

[21] Bootstrapping is based on 1000 replications.

The extent of imperfect proxy bias when using FSM status as a control variable is also context-specific. The third part of the analysis assesses the extent of bias when estimating the effects of school type and Special Educational Needs (SEN) on KS2 attainment. Following the conceptual framework, four OLS regressions are estimated:

(1) A "true" regression of KS2 attainment on covariates **x** and SES variables;
(2) An omitted variables regression of KS2 attainment on covariates **x**;
(3) A proxy variable regression of KS2 attainment on covariates **x** and FSM status; and
(4) A regression of KS2 attainment on covariates **x**, SES variables and FSM status.

where the covariates **x** include school type, SEN and Key Stage 1 attainment.[22]

Regressions (1)-(3) mirror equations (1)-(3) in section 2.2. Equation (4) assesses the redundancy of FSM status in the "true" model. The estimated "omitted variables bias" on school type is the difference in the school type parameter in the omitted variables and "true" regressions. The estimated "imperfect proxy bias" on school type is the difference in the school type parameter in the proxy variable and "true" regressions. Biases on SEN are estimated equivalently. Again bootstrapping is used to estimate the standard error of these biases and test if they are statistically significant.

If FSM status is redundant in the "true" model, then these are consistent estimates of the "true" biases (as defined in section 2.2). However, if it is not redundant, then the estimated biases themselves suffer from attenuation bias.[23]

[22] The other covariates are age, gender, ethnicity, English as a First Language and Key Stage 2 assessment year. Standard errors are corrected for the clustering of units in schools and heteroskedasticity.

[23] Assuming parameters have "natural" signs. See appendix 2.

5. Data

The main dataset is the Avon Longitudinal Study of Parents and Children (ALSPAC). This is our source of SES measures. The National Pupil Database (NPD), matched into ALSPAC, is our source of FSM status, Key Stage 1 and 2 attainment, school type and Special Educational Needs. Section 5.1 describes ALSPAC and NPD, section 5.2 describes the principal measures, and section 5.3 describes sample selection.

Datasets: ALSPAC and NPD

All pregnant women, resident in Avon, with an expected date of delivery between the 1st April 1991 and 31st December 1992 were eligible for ALSPAC.[24] An estimated 85-90% of the eligible population enrolled in the study (O'Connor et al., 1999, p779). The core sample consists of 14,049 live births available for analysis.[25]

NPD contains administrative data on all pupils in state schools in England, collected by the Department for Education and Skills. Key Stage 1 (KS1) and Key Stage 2 (KS2) data were matched to 87% of the core ALSPAC sample. Pupil Level Annual School Census (PLASC) 2002 data were matched to 84% of the core sample.

Table 1 assesses the extent of non-random matching, reporting the difference in means of (standardised) mother's age at birth in the matched and unmatched samples. Mother's age at birth is available for all children and is positively correlated with KS2 attainment, for those with matched data (Pearson's correlation coefficient is 0.20-0.21). For KS1 and KS2 data, the

[24] The study area is well-defined, consisting of that part of the county of Avon that was also within the then South West Regional Health Authority. It therefore excluded Bath and district.

[25] A comparison of the ALSPAC sample completing questionnaires 8 months after birth with the population of mothers with infants under 1 year of age resident in Avon at the time of the 1991 Census, found that the ALSPAC sample had more owner-occupiers (79% vs. 69%), more households with a car (91% vs. 84%), more married couples (79% vs. 72%) and fewer mothers from minority ethnic groups (2.2% vs. 4.1%), but more households with one or more persons per room (34% vs. 26%). Reported on the ALSPAC website: http://www.alspac.bris.ac.uk (Accessed: 1 September 2006).

extent of non-random matching is quite low. For PLASC data, non-random matching is more severe: the difference in means is 0.17 standard deviation units. The mean of mother's age is less in the matched samples, perhaps because these exclude pupils in private schools (whose mothers are older, on average).

Principal measures

The following measures of SES are constructed from ALSPAC:

Family income at age 4

Mothers are asked, "On average, about how much is the take home family income each week (include social security benefits etc.)?" Responses are banded "less than £100", "£100-199", "£200-299", "£300-£399" and "£400 or more".

Mother's, partner's and family employment, and one-parent family status at age 4½

One-parent family status is constructed from mothers' reports of having a "partner".[26] Mother's and partner's employment are coded "full-time", "part-time" or "not employed". Family employment is coded "no employment", "one part-time", "one full-time or two part-time", "one full-time and one part-time" and "two full-time".

Mother's and partner's highest educational qualification in pregnancy

Mother's and partner's highest educational qualifications are constructed from mothers' and partners' reports of their own and their partner's qualifications.[27] There are five categories: "CSE/no qualifications", "vocational", "O-level", "A-level" and "degree".[28]

[26] 97% of partners at age 4 "live-in" (and IS and IB-JSA eligibility rules refer to live-in partners). Furthermore, 94% of partners at age 4 are biological parents (mostly fathers) of the child.

[27] The variables are constructed from self-reports unless these are missing.

[28] Thirteen qualifications and "no qualifications" are grouped, labelled and ordered as follows: {"No qualifications", "CSE or GCSE (D, E, F or G)"} are labelled "CSE/no qualifications" and ordered first; {"Qualifications in shorthand/typing/or other skills, e.g., hairdressing", "Apprenticeship", "City & Guilds intermediate technical", "Other"} are labelled "vocational" and ordered second; {"O-level or GCSE (A, B or C)"} are labelled "O-level" and ordered third; {"A-level", "State enrolled nurse", "State registered nurse", "City & Guilds final technical", "City & Guilds full technical", "Teaching qualification"} are labelled "A-level" and ordered fourth; and {"University degree"} is labelled "degree" and ordered fifth.

Mother's and partner's social class in pregnancy

Mother's and partner's social class are constructed using the 1991 Office of Population Censuses and Surveys "Standard Occupational Classification".[29] There are six classes: I Professional, II Managerial and Technical, IIINM Skilled Non-Manual, IIIM Skilled Manual, IV Partly Skilled and V Unskilled.

Table 2 assesses the extent of non-random non-response for these SES measures. Response rates vary from 89% for mother's education to 62% for family income. Non-response is non-random. The means of mothers' ages in samples with each SES measure are 0.34-0.45 standard deviation units greater than those in samples without them. These differences are large and statistically significant.

The following measures are constructed from NPD:

FSM status at age 9-11

FSM status is extracted from PLASC 2002 data and is available for all but one of the children with PLASC 2002 data. 11.8% of these children claim FSM. The children are aged 9-11 in PLASC 2002.[30] This is much older than when the SES variables are observed in pregnancy, or aged 4 or 4½. This limitation of the data is discussed later.

Key Stage 2 attainment

At the end of KS2, children take tests in English, maths and science. The tests are externally marked and award 0-100 marks in each subject. Normalised, KS2 attainment measures in each subject are constructed as follows. The 1.1% to 2.4% of children working at levels 1 and 2, and thus not entered for the tests, are scored zero. Those entered for the tests are scored their mark on the tests. Children are then ranked, randomly splitting ties, and ranks are converted to z scores.[31] The 1.5% of children "disapplied" or "absent" from the tests are coded missing and dropped from estimation.[32]

[29] This is also called the "Registrar General's Social Classes". The classification has it limitations (Rose, 1995; Rose and O'Reilly, 1998). However, this is the only classification currently available in ALSPAC.

[30] The age variation occurs because the children are in three school cohorts.

[31] This is a relatively common approach in research on educational test scores (Goldstein 2003, p31).

[32] Children assessed by teachers to be working at level 6 in a subject are also entered for an extension test. However, no children in the estimation samples sat these tests.

Key Stage 1 attainment

In the final year of KS1, unless assessments are "disapplied", children take a combination of tasks and tests in three subjects: reading, writing/spelling and maths.[33] In each subject, a categorical measure of attainment is constructed based on the complete set of outcomes from the task and tests (with the exception of outcomes with a small number of cases). This results in measures of reading and writing/spelling each with 14 categories, and a maths measure with seven categories.[34]

School type

School type at the time of KS2 examinations is coded "community", "voluntary-aided", "voluntary-controlled" and "other".

Special Educational Needs

Special Educational Needs is coded "statement" and "no statement".[35]

Other covariates

Other covariates are: gender, age at KS2 examinations, KS2 assessment year, ethnicity and English as a First Language.

Sample selection

Analysis of the joint distributions of FSM status and each SES variable is restricted to children with complete data on FSM status and the relevant SES variable. In addition, analysis of the joint distributions of FSM status, and partner's employment, education and

[33] Only 0.3-0.4% of the core sample with KS1 data are "disapplied". Assessments are marked by schools.

[34] KS1 reading categories: W, 1, (task=2A, tests=2A), (2A, 2B), (2A, 2C), (2B, 2A), (2B, 2B), (2B, 2C), (2C, 2A), (2C, 2B), (2C, 2C), (2C, L), 3/4+, and "disapplied". KS1 writing/spelling categories: W, (task=1,test=X), (1,L), (2C, L), (2C, 2), (2C, 3), (2B, L), (2B, 2), (2B, 3), (2A, 2), (2A, 3), (3,2), (3/4+,3), and "disapplied". KS1 maths categories: W, 1, 2C, 2B, 2A, 3/4+, and "disapplied".

[35] In PLASC 2002 data, schools used either old or new SEN codes. The vast majority of schools used the old codes in the matched sample.

social class is restricted to children with a mother with a partner at age 4½.[36] Sample sizes vary from 10,505 to 6,606 children (75% to 47% of the core sample).

Table 3 assesses the extent of non-random sample selection. Sample selection is strongly non-random. In particular, the means of mothers' ages in the selected samples are 0.08-0.35 standard deviation units greater than those in the non-selected samples. In addition, only 4.3-9.7% of children claim FSM in the selected samples, compared to 11.8% of those with FSM data.

Analysis of differences in KS2 attainment by family income is restricted to children with complete data on KS2 attainment, family income, FSM status and gender. This sample is 7,214 children (51% of the core sample). The mean of mother's age in the selected sample is 0.20 standard deviation units greater than in the non-selected sample. Moreover, only 7.9% of children claim FSM in the selected sample.

Analysis of the effects of school type and Special Educational Needs on KS2 attainment is restricted to children with complete data on KS2 attainment, school type, Special Educational Needs, FSM status, KS1 attainment and the "other covariates".[37] This sample is 11,130 children (79% of the core sample). The mean of mother's age in the selected sample is 0.12 standard deviation units *less* than in the non-selected sample, and 11.2% of children claim FSM in the selected sample, similar to the 11.8% of children with FSM data.

6. FSM Eligibility Rules and Nationally-Representative Claimant Data

Before presenting the main results, this section describes the FSM measure and FSM eligibility rules, presents nationally-representative data on those eligible for FSM, discusses factors affecting decisions to claim FSM, and describes continuities and discontinuities in FSM status over time.

[36] A potential problem is that the partner when the study child is age 4½ could differ from the partner during pregnancy when partner's education and social class is measured. This is not known. However, 94% of partners when the study child is age 4 are biological parents of the study child.

[37] Non-response on the categorical SES variables is addressed with a missing category.

The FSM measure: eligible for and claiming FSM

DfES's PLASC Completion Notes for primary schools in 2006 states that:

*Pupils should be recorded as eligible **(Y) ONLY** if a **claim** for free school meals has been made by them or on their behalf by parents **and** either*

*(a) the relevant authority has confirmed the eligibility and the free school meal is currently provided for them, **or***

(b) the school or LEA have seen the necessary documentation (for example, an Income Support order book) that supports their eligibility, and the administration of the free meal is to follow as a matter as process.

*Conversely, if pupils are in receipt of a free meal but there is confirmation that they are no longer eligible and entitlement will be revoked, code **N** should be applied.*

DfES (2006, p12; original emphasis)

Similar wording is used in PLASC Completion Notes for primary schools and secondary schools in 2002-2005. Critically, the "FSM eligibility" measure is not, therefore, a measure of eligibility but of being *eligible for and claiming* FSM.

FSM eligibility

FSM eligibility is itself contingent upon being *eligible for and claiming* other benefits. FSM eligibility rules have changed over time. In 2001/02, the year relevant to our main analysis, children in families claiming Income Support (IS) or Income-Based Jobseekers Allowance (IB-JSA) were eligible for FSM.[38] Section 6.3 summarises eligibility rules for IS and IB-JSA

[38] In 2005/06, by contrast, children of families receiving the following benefits were eligible for FSM: IS or IB-JSA; Support under part VI of the Immigration and Asylum Act 1999; Child Tax Credit provided they do not

in that year. Section 6.4 presents nationally-representative data on IS and IB-JSA claimants in May 2002.

IS and IB-JSA eligibility in 2001/02[39]

Income Support (IS) claims are made and assessed on a 'benefit unit' basis. A benefit unit consists of a claimant plus any partner and dependent children. The claimant must be aged 16 or over. They must not be working 16 hours or more a week or have a partner working 24 hours or more per week. In addition, the claimant must not be required to be available for employment. The benefit unit's income ('resources') must be below their needs ('applicable amounts'). Applicable amounts consist of a 'personal allowance' which depends upon the age of the claimant and the presence and age of a partner; additions for any dependents; 'premiums' which provide additional allowances in recognition of special needs such as old age or disability; and certain types of housing costs. Finally, their capital assets must be less than £8,000 for claimants aged under 60.

To be entitled to Job Seekers Allowance (JSA) claimants must be available to work for at least 40 hours per week; be actively seeking work; be capable of work; not be in relevant education; be out of work or working on average for less than 16 hours per week, and, in the case of Income-Based Job Seekers Allowance (IB-JSA), any partner must be working less than 24 hours a week on average; and be under pensionable age. Claimants who have paid sufficient National Insurance contributions get Contribution-Based JSA, at a personal rate for up to six months. Those who do not qualify for, or whose needs are not met by, Contribution-Based JSA qualify for IB-JSA if, like IS, their 'resources' are below their 'applicable amounts' and their capital assets are less than £8,000 for claimants aged under 60.[40]

In short, to be eligible for IS or IB-JSA in 2001/02, children must be in a "benefit unit" without a member working more than 24 hours per week, with a low income (defined relative to needs), and limited capital assets.

---

receive Working Tax Credit and have an annual income from 6 April 2005 (as assessed by The Inland Revenue) which does not exceed £13,480; and Guarantee element of State Pension Credit.

[39] This section draws heavily on DWP (2002a, b).

[40] 16- and 17-year-olds are normally unable to claim IS or JSA.

IS and IB-JSA claimants in May 2002

Not everyone who is eligible for IS or IB-JSA claims them. DWP (2004, p15-16) estimated that 93-99% of eligible non-pensioners with children claimed IS in 2001/02. Estimated take-up rates were higher for lone parents (94-100%) than for couples with children (85-94%). Estimated take-up rates for IB-JSA in the same year were lower; 70-81% for couples with children (DWP 2004, p44). For both benefits, take-up rates were higher for those with lower incomes.[41]

DWP conducts quarterly sample surveys of the populations of IS and IB-JSA claimants in Great Britain. The remainder of this section reports findings from these surveys in May 2002, as reported in DWP (2002a, b). In May 2002, 92% of dependents of claimants of IS and IB-JSA were dependants of IS claimants; 8% were dependants of IB-JSA claimants.

Table 4 reports the number of dependents of IS claimants by 'statistical group' and age of dependent. 72% of dependents have 'lone parent' and 22% have 'disabled' claimants. The proportion of dependents with 'lone parent' claimants was greater for those of primary school age (75%) than for those of secondary school age (67%). However, these figures understate the proportion of dependents with single parent claimants because some 'aged 60 or over' and 'disabled' claimants are single parents.[42] In particular, 83% of IS claimants with dependents were single and 17% were in a couple. The majority of 'lone parent' claimants are aged 25-59 (80%) and have one or two dependents (77%).

Although some claimants have other income sources, IS payments are likely to be the main source of income for most claimants. In May 2002, the average IS payment was £106 per week for single claimants with dependents and £129 for couples with dependents. IS

[41] These are caseload take-up estimates. "Caseload take-up compares the number of benefit recipients – averaged over the year – with the number who would be receiving if everyone took up their entitlement for the full period of their entitlement" (DWP 2004, p3). The estimated ranges are 95% confidence intervals.

[42] Claimants are allocated to statistical groups based firstly upon their entitlement to the pensioner or disability premiums, 'Aged 60 or over' and 'Disabled' groups, respectively, and then to the 'Lone parent' group if they are single with dependents and not already classified as pensioner or disabled.

payments were greater for claimants with more dependents and older claimants. For example, the average IS payment was £85 for 'lone parents' with one dependent and £132 for those with three dependents. Table 5 reports the duration of IS claims. While 17% of all claimants and 21% of 'lone parent' claimants received IS for one year or less, 44% of all claimants and 34% of 'lone parent' claimants received IS for five or more years.

Only 8% of dependents of IB-JSA claimants have single parents. Moreover, only 3% of dependents aged 5-15 have single parent claimants. As with IS, the majority of IB-JSA claimants with dependents are aged 25-49 (80%) and have one or two dependents (71%). IB-JSA payments are likely to be the main source of income for most claimants. In May 2002, the average IS payment was £123 per week for claimants with dependents. Finally, as reported in table 6, IB-JSA claims were of much shorter duration than IS claims; 76% of all claimants received IB-JSA for one year or less.

In summary, pupils in families claiming IS or IB-JSA were eligible for FSM in 2001/02. The vast majority of these were in families with one parent, aged 25-59, and one or two children. Given IS and IB-JSA eligibility rules, these pupils should be in families without a parent in full-time employment, with low incomes (defined relative to needs) and limited capital assets.

Claiming FSM

Croxford (2000, p318-319) argued,

There is much speculation, but no research evidence, concerning the extent to which parents' decisions [to claim free school meals] are influenced by the perceived stigma attached to claiming free school meals, or the dietary requirements of some groups, or the enthusiasm with which schools and local authorities encourage parents to seek entitlement.

Storey and Chamberlin (2001, p2-3) supported some of these contentions, however.[43] They found that one third of pupils and two fifths of parents said embarrassment or fear of being

[43] The study visited 7 secondary, 2 middle and 4 primary schools in 7 LEAs. 250 pupils were interviewed, 450 completed questionnaires, and 50 parents were interviewed.

teased put them off claiming FSM. Parents wanted to protect their children from being seen as different. Teasing and bulling were more common in schools with few FSM pupils. In most schools, FSM pupils continued to be identifiable, for example, because they had to give their names or tokens when collecting their meal.[44] Other problems that caused parents not to claim FSM included the fact that their children often had to sit separately from their friends, and the quality of the food on offer was not high and the options often unhealthy.[45] There are therefore likely to be systematic reasons for non-claiming.

FSM persistence

Finally, the dynamics of FSM status over time are examined. Table 7 presents data for the cohort of pupils entering reception year in 1997/98.[46] 18% of this cohort claimed FSM in January 2004 (in year 6). However, 86% of those claiming FSM in 2003 also claimed FSM in 2004. An even large proportion of non-FSM pupils in 2003 were also non-FSM pupils in 2004 (97%). Furthermore, 13% of this cohort were FSM in 2002, 2003 and 2004, and 77% were non-FSM in all three years; only 10% changed FSM status over this period. These relatively strong continuities in FSM status over time are consistent with the evidence on the duration of IS and IB-JSA claims reported earlier.

7. Results

The joint distributions of FSM status and the SES variables

Figure 1 presents the joint distributions of FSM status and our nine measures of SES.[47] Table 8 reports the binary indicator of each SES measure which FSM status proxies "best", along with the associated probabilities of false positives and negatives.

[44] Recall that the research was undertaken more than five years ago and this may no longer be the case.

[45] Another possibility is that parents are less likely to claim FSM the lower the proportion of pupils in the school claiming FSM.

[46] The pattern of dynamics is very similar for the cohorts entering reception year in 1995/96 and 1996/97.

[47] Appendix table 1 reports the distributions in tabular form.

For family income, FSM status proxies "best" incomes below £200 per week, the bottom income quartile, not the bottom decile, even though less than 10% of children claim FSM in this sample. However, 22% of FSM children have incomes above £200 per week (false positives), and 20% of non-FSM children have incomes below £200 per week (false negatives).

For family employment, FSM status proxies "best" workless families and those with only one part-time worker.[48] While only 8% of non-FSM children are in workless families and those with one part-time worker (false negatives), 43% of FSM children are in families with one or more full-time workers, or two part-time workers (false positives).

FSM status is a far from perfect proxy of one parenthood. In particular, 72% of FSM children have two-parents (false positives). FSM status is also a far from perfect proxy of mothers' and partners' education and social class. For each of these measures, the probabilities of false positives and/or false negatives is high.[49]

Anomalies

There are three anomalies between these data, and FSM eligibility rules and IS/IB-JSA claimant data. First, 43% of FSM children are in families with at least one full-time worker, even though these families should not be eligible for IS or IB-JSA, and hence FSM. Second, only 28% of FSM children are in one-parent families. This is much lower than one would expect from IS/IB-JSA claimant data, which suggests that over 75% of primary school age dependents of IS claimants are in one-parent families. Third, only 42% of children in workless families *and* with incomes below £100 per week claim FSM.

One explanation for these anomalies is that the SES variables are observed in pregnancy or between ages 4-4½, but FSM status is observed between ages 9-11. The SES variables, and

[48] As defined here, "workless families" include one-parent families where this parent is not working and two-parent families where neither parent is working.

[49] We examined the joint distribution of FSM status and partner's education, restricting the sample to children with a mother with a partner in pregnancy (rather than at age 4½). This sample suffers less from attrition than the one in figure 1. We found that the joint distributions were similar in both samples. The same was true of the joint distributions of FSM status and partner's social class.

hence FSM status, inevitably change over time. Section 6 reported that 10% of children change FSM status over a 2 year period at the end of primary school/start of secondary school, and only 44% of IS claimants receive the benefit continuously for 5 or more years. Supporting this explanation, 61% of children in continuously workless families (i.e., in workless families in pregnancy, and at 33, 47 and 54 months) claim FSM, compared to 42% of children in workless families at 54 months.

A second explanation for the first two anomalies is that some families misreport their employment, income and/or partnership status when claiming benefits. A second explanation for the third anomaly is that workless families with low incomes are eligible for IS/IB-JSA and FSM, but not claiming them. Section 6 reported take-up rates of between 93-99% for IS and 70-81% for IB-JSA. We are unaware of any study estimating the proportion of children eligible for FSM who claim them.

Estimating differences in Key Stage 2 attainment by family income

As stated, the extent of imperfect proxy bias when using FSM status as the variable of interest is context-specific. We assess this bias when estimating differences in KS2 attainment by family income (using FSM status to proxy incomes below £200 per week).

Table 9 reports the results. For example, the difference in the average KS2 English attainment of girls with family incomes below and above £200 per week is 0.50 standard deviation units. However, the difference in the average KS2 English attainment of FSM and non-FSM girls is 0.67 standard deviation units. In this case, the imperfect proxy bias is 0.17 standard deviation units, 34% of the "true" difference. With one exception, the imperfect proxy bias is 30-40% of the "true" difference and is significant at the 1% level. In this context, imperfect proxy bias is quite large.

Estimating school type and special educational needs effects on KS2 attainment

The extent of imperfect proxy bias when using FSM status as a control variable is model-specific. We assess this bias when estimating the effects of school type and Special Educational Needs on KS2 attainment, using FSM status to proxy eight SES measures.

Table 10 reports the results. Column 1 is the "true" regression of KS2 attainment on school type, Special Educational Needs, SES measures, KS1 attainment and the other covariates.[50] Column 2 omits the SES measures. We refer to this as the omitted variables regression. Column 3 omits the SES measures but adds FSM status. We refer to this as the proxy variable regression.

Consider first the effect of voluntary-aided schools, relative to community schools. In English, the effect is positive and significant in both the "true" and proxy variable regressions. However, in maths and science, the effect is significant in the proxy variable regression, but is *insignificant* in the "true" regression. In each subject, the difference in the parameter in the "true" and proxy variable regressions, the imperfect proxy bias, is large, 30-64% of the "true" parameter, and significant. More importantly, the imperfect proxy bias is 90% of the omitted variables bias, the difference in the parameters in the "true" and omitted variables regressions. In other words, adding FSM status to the regression reduces the omitted variables bias by only 10%.

Consider next the effect of voluntary-controlled schools, relative to community schools. In English, the effect is insignificant in the "true" and proxy variable regressions. In science, the effect is negative and significant in both regressions. However, in maths, the effect is negative and insignificant in the proxy variable regression, but is negative and *significant* in the "true" regression. In each subject, the imperfect proxy bias is significant and 75-80% of the omitted variables bias.

[50] The other covariates are age, gender, ethnicity, English as a First Language and KS2 assessment year.

Finally, consider the effect of Special Educational Needs. For every subject, the effect is negative and significant in both the "true" and proxy variable regressions. Nevertheless, the imperfect proxy bias is significant and 75-80% of the omitted variables bias.

In summary, adding FSM status to the models without measures of SES, reduces the omitted variables bias by only 10-25%. Furthermore, because FSM status is not redundant when added to the "true" model (column 4 of table 10), the estimated omitted variables and imperfect proxy biases could suffer themselves from attenuation bias (see appendix 2). In short, FSM status has limited value in this context.

8. Concluding Remarks

Future research should evaluate the FSM measure in other contexts and datasets, and the use of various small area data matched to children's home postcodes as proxies for measures of SES. In the meantime, researchers should be cautious in drawing inferences from research reliant on the FSM measure. When used as the variable of interest, FSM status is an imperfect proxy of low income or "workless" families, or one-parenthood. In the context of estimating differences in educational attainment by family income, imperfect proxy bias is quite large. When used as a control variable in an OLS regression, FSM status reduces omitted variables bias to a moderate extent only. In other words, if omitted variables bias is a concern, then the inclusion of FSM status in the model should do little to diminish this concern.

Figure 1 Distributions of SES variables Conditional on FSM Status

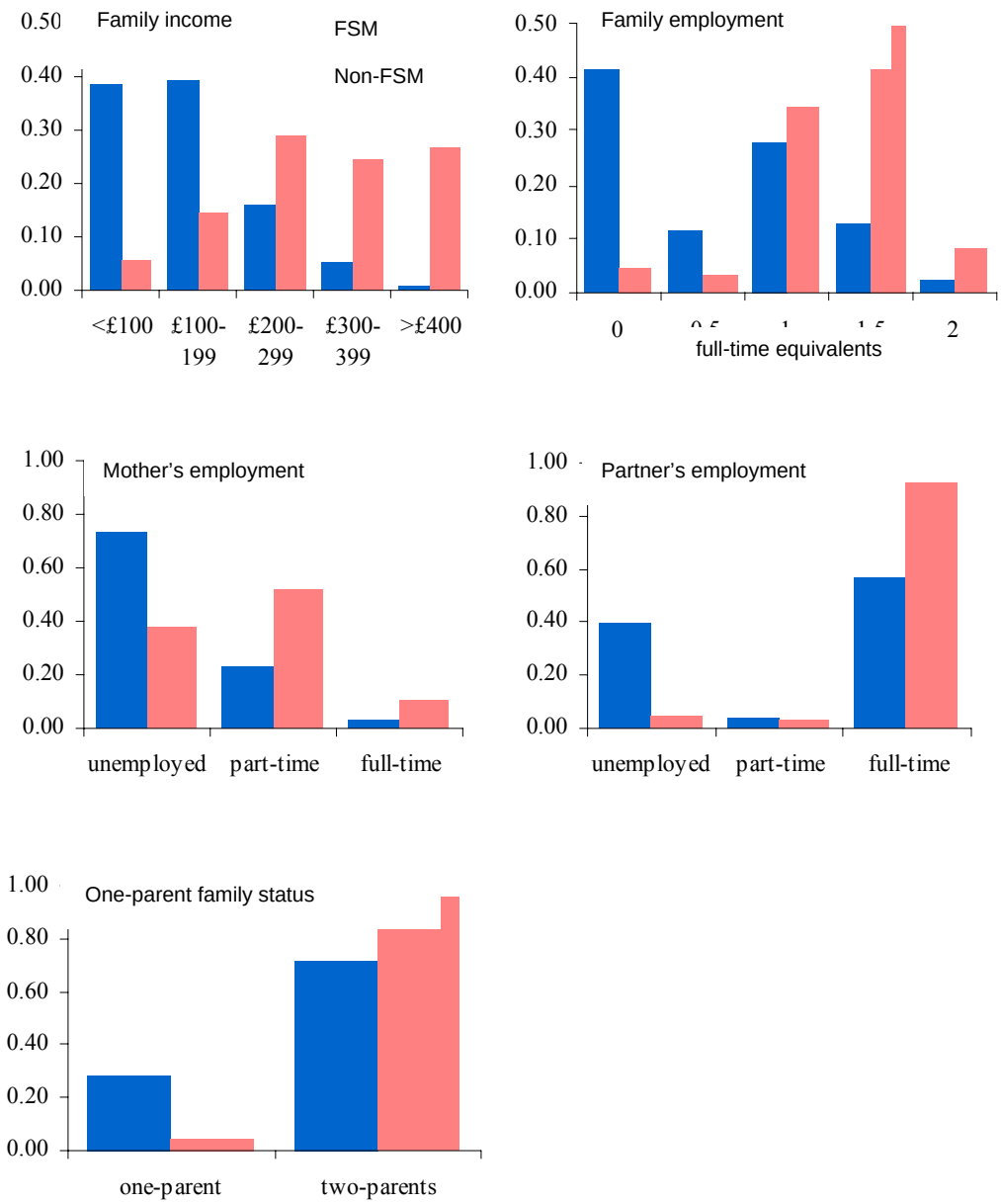

Figure 1 (continued) Distributions of SES variables Conditional on FSM Status

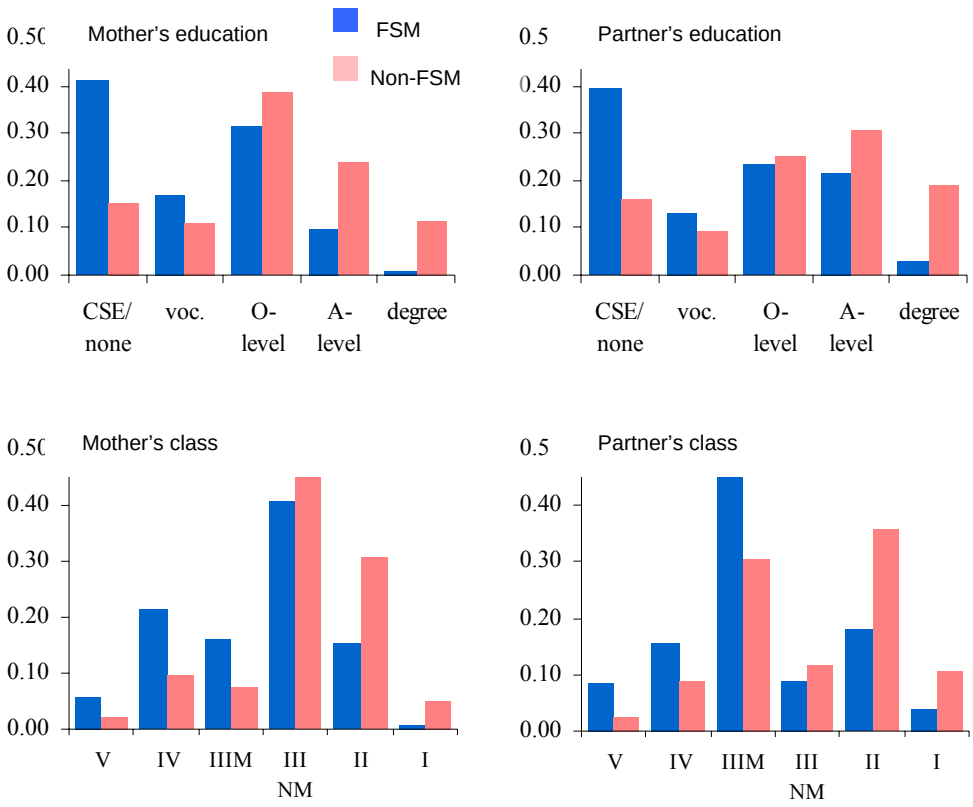

Table 1 NPD Matching

Selected sample	**Sample size** Response rate % (1)	**Differences in sample means** (standard error) Mother's age
PLASC 2002 data	11,844 (84.3)	-0.171*** (0.024)
KS2 data	12,234 (87.1)	-0.036 (0.027)
KS1 data	12,158 (86.5)	-0.081*** (0.026)

Response rate (1)=sample size/14,049.
Difference in sample means *equals* selected sample mean *minus* non-selected sample mean.
Standardised mother's age at birth of study child. Combined sample of 14,049 children.
Two-tailed, two-sample t-test with unequal variances.
*, ** and *** indicate significance at the 10%, 5% and 1% levels, respectively.

Table 2 SES Measures: Response and Non-response

Selected sample	**Sample size** Response rate % (1) [2] {3}	**Differences in sample means** (standard error) Mother's age
Family income/47m	8,645 (61.5) [68.5] {89.9}	0.341*** (0.018)
Family employment/54m	8,957 (63.8) [69.2] {92.2}	0.445*** (0.018)
Mother employment/54m	9,327 (66.4) [69.2] {96.0}	0.427*** (0.018)
Partner employment/54m	9,084 (64.7) [69.2] {93.5}	0.442*** (0.018)
One-parent family status/54m	9,084 (64.7) [69.2] {93.5}	0.442*** (0.018)
Mother education/pregnancy	12,458 (88.7)	0.433*** (0.029)
Partner education/pregnancy	12,069 (85.9)	0.452*** (0.026)
Mother social class/pregnancy	10,090 (71.8) [89.2] {80.5}	0.339*** (0.020)
Partner social class/pregnancy	11,156 (79.4) [89.2] {89.0}	0.450*** (0.022)

Response rates: (1)=sample size/14,049, [2]=wave response rate, {3}=item response rate.
Difference in sample means *equals* selected sample mean *minus* non-selected sample mean.
Standardised mother's age at birth of study child. Combined sample of 14,049 children.
Two-tailed, two-sample t-test with unequal variances.
*, ** and *** indicate significance at the 10%, 5% and 1% levels, respectively.

Table 3 Sample Selection

Selected sample	**Sample size** Response rate % (1)	**Difference in sample means** (standard error) Mother's age	**Sample mean** FSM status
Joint distributions			
Family income & FSM	7,473 (53.2)	0.208*** (0.017)	0.081
Family employment & FSM	7,708 (54.9)	0.275*** (0.017)	0.066
Mother employment & FSM	8,039 (57.2)	0.249*** (0.017)	0.073
Partner employment & FSM	7,355 (52.4)	0.306*** (0.017)	0.052
One-parent family status & FSM	7,821 (55.7)	0.268*** (0.017)	0.068
Mother education & FSM	10,505 (74.8)	0.078*** (0.021)	0.097
Partner education & FSM	6,930 (49.3)	0.343*** (0.017)	0.045
Mother social class & FSM	8,505 (60.5)	0.161*** (0.018)	0.073
Partner social class & FSM	6,606 (47.0)	0.347*** (0.016)	0.043
Estimation samples			
Differences in attainment sample	7,214 (51.3)	0.202*** (0.017)	0.079
School type & SEN effects sample	11,130 (79.2)	-0.117*** (0.022)	0.112

Response rate (1)=sample size/14,049.
Difference in sample means *equals* selected sample mean *minus* non-selected sample mean.
Standardised mother's age at birth of study child. Combined sample of 14,049 children.
Two-tailed, two-sample t-test with unequal variances.
*, ** and *** indicate significance at the 10%, 5% and 1% levels, respectively.
FSM status in 2002: claiming=1, not claiming=0.
Partner's employment, partner's education and partner's social class cross-classifications restricted to children with a mother with a partner at 54 months.

Table 4 Characteristics of Income Support Claimants (May 2002)

Percentage of dependents of IS claimants by statistical group and age of dependent					
	< 5	5–10	11–15	16+	*All*
Aged 60 or over	0	1	3	9	*2*
Disabled	16	20	26	38	*22*
Lone parent	79	75	67	50	*72*
Other	4	4	4	4	*4*
Total	*100*	*100*	*100*	*100*	*100*

Own calculations from DWP (2002a).

Table 5 Duration of Income Support Claims (May 2002)

Duration	All Claimants (%)	Lone Parents (%)
Under 3 months	5	5
3 to under 6 months	4	6
6 to under 12 months	8	10
1 to under 2 years	15	16
2 to under 3 years	10	12
3 to under 4 years	8	10
4 to under 5 years	6	7
5 years or over	44	34
Total	*100*	*100*

Source: DWP (2002a) Tables 3.1 and 3.2.

Table 6 Duration of Income Based-Job Seekers Allowance Claims (May 2002)

Duration	All Claimants (%)
Up to 2 weeks	6
2 to under 6 weeks	12
6 to under 13 weeks	16
13 to under 26 weeks	21
26 to under 39 weeks	14
39 weeks to under 1 year	8
1 to under 2 years	13
2 years or over	12
Total	*100*

Own calculations from DWP (2002b) Table 2.2.

Table 7 FSM Status Dynamics

FSM Status in 2002	FSM Status in 2003	Probability FSM Status in 2004 is "yes" (%)
-	-	18
-	yes	86
-	no	3
yes	yes	87
no	yes	75
yes	no	23
no	no	2

PLASC 2002, 2003 and 2004 data. Cohort entering reception year in 1997/98. Own calculations. 573,676 pupils with complete data.

Table 8 Imperfect Binary Proxy: False Positives and Negatives

SES measure	Low SES	False positives	False negatives
Family income	{< £200 per week}	0.22	0.20
Family employment	{0 employed, 1 part-time}	0.43	0.08
Mother's employment	{unemployed}	0.26	0.37
Partner's employment	{unemployed, part-time}	0.57	0.07
One-parent family	{one-parent}	0.72	0.04
Mother's education	{Vocational, CSE/none}	0.42	0.26
Partner's education	{Vocational, CSE/none}	0.48	0.26
Mother's class	{IIIM, IV, V}	0.57	0.19
Partner's class	{IIIM, IV, V}	0.31	0.42

Family income: N=7473, P(FSM=yes)=8.1%, P(SES=low)=24.6%.
Family employment: N=7708, P(FSM=yes)=6.6%, P(SES=low)=11.0%.
One-parent family status: N=7821, P(FSM=yes)=6.8%, P(SES=low)=6.0%.
Mother's employment: N=8039, P(FSM=yes)=7.3%, P(SES=low)=40.1%.
Partner's employment: N=7355, P(FSM=yes)=5.2%, P(SES=low)=6.6%.
Mother's education: N=10505, P(FSM=yes)=9.7%, P(SES=low)=29.3%.
Partner's education: N=6930, P(FSM=yes)=4.5%, P(SES=low)=26.8%.
Mother's class: N=8505, P(FSM=yes)=7.3%, P(SES=low)=21.0%.
Partner's class: N=6606, P(FSM=yes)=4.3%, P(SES=low)=42.9%.
Partner's employment, partner's education and partner's social class samples restricted to children with a mother with a partner at 54 months.
False positives: Prob(SES>low|FSM=yes).
False negatives: Prob(SES=low|FSM=no).

Table 9 FSM and Low-Income Attainment Gaps: Imperfect Proxy Bias

	Differences in KS2 attainment by:			
	Low-income status	FSM status	**Imperfect proxy bias**	
	Est. (Std. err.)	Est. (Std. err.)	Est. (Std. err.)	Percent
English				
Girls	-0.504*** (0.042)	-0.673*** (0.056)	-0.169*** (0.053)	-33.5
Boys	-0.509*** (0.040)	-0.703*** (0.069)	-0.194*** (0.059)	-38.1
Maths				
Girls	-0.516*** (0.037)	-0.592*** (0.054)	-0.076 (0.049)	-14.7
Boys	-0.491*** (0.042)	-0.692*** (0.072)	-0.201*** (0.061)	-40.9
Science				
Girls	-0.493*** (0.042)	-0.648*** (0.058)	-0.155*** (0.053)	-31.4
Boys	-0.476*** (0.044)	-0.669*** (0.067)	-0.194*** (0.061)	-40.7

N=7214 children. Girls=3573, Boys 3641.
Girls & FSM=316, Girls & non-FSM=3257 (8.8% FSM).
Girls & low-income=864, Girls & higher-income=2709 (24.2% low-income).
Boys & FSM=253, Boys & non-FSM=3388 (7.0% FSM).
Boys & low-income=881, Boys & higher-income=2760 (24.2% low-income).
Low-income status equals 1 if <£200 per week, 0 otherwise.
Differences in KS2 attainment estimated in OLS regression of KS2 attainment on low-income status (or FSM status) and KS2 assessment year.
Standard errors on differences in KS2 attainment are robust and corrected for clustering at school-level.
Standard errors on imperfect proxy bias are bootstrap standard errors (1000 replications).
*, ** and *** indicate significance at the 10%, 5% and 1% levels.
Imperfect proxy bias = estimated FSM gap – estimated low-income gap.
Imperfect proxy bias (%) = (imperfect proxy bias/|estimated low-income gap|)*100.

Table 10 SEN and School Type Effects: Omitted Variables and Imperfect Proxy Biases

	(1) "True" model	(2) Omitted variables model	(3) Proxy variable model	(4) "True" model & FSM	(5) Omitted variables bias =(2)-(1)	(6) Imperfect proxy bias =(3)-(1)
	Est. (Std. err.)	Est. (Std. err.)	Est. (Std. err.)	Est. (Std. err.)	Est. [%] (Std. err.)	Est. [%] (Std. err.)
English						
Voluntary-aided	0.107*** (0.040)	0.142*** (0.040)	0.139*** (0.040)	0.106*** (0.040)	0.036*** (0.005) [33.2]	0.032*** (0.005) [29.8]
Voluntary-controlled	0.015 (0.028)	0.045 (0.031)	0.037 (0.031)	0.012 (0.028)	0.030*** (0.004) [201]	0.023*** (0.004) [154]
Statement of SEN	-0.272*** (0.059)	-0.232*** (0.059)	-0.242*** (0.059)	-0.275*** (0.059)	0.040*** (0.011) [14.8]	0.030*** (0.011) [11.2]
FSM status	- -	- -	-0.134*** (0.020)	-0.078*** (0.020)		
Maths						
Voluntary-aided	0.057 (0.043)	0.090** (0.043)	0.087** (0.042)	0.056 (0.043)	0.033*** (0.005) [58.1]	0.030*** (0.005) [52.8]
Voluntary-controlled	-0.068** (0.028)	-0.038 (0.031)	-0.044 (0.031)	-0.070** (0.028)	0.030*** (0.004) [43.9]	0.024*** (0.004) [35.3]
Statement of SEN	-0.267*** (0.057)	-0.223*** (0.058)	-0.231*** (0.058)	-0.269*** (0.057)	0.044*** (0.011) [16.4]	0.036*** (0.010) [13.3]
FSM status	- -	- -	-0.112*** (0.020)	-0.051*** (0.019)		
SES measures included	yes	no	no	yes		

Table 10 (continued) SEN and School Type Effects: Omitted Variables and Imperfect Proxy Biases

	(1) "True" model	(2) Omitted variables model	(3) Proxy variable model	(4) "True" model & FSM	(5) Omitted variables bias =(2)-(1)	(6) Imperfect proxy bias =(3)-(1)
	Est. (Std. err.)	Est. (Std. err.)	Est. (Std. err.)	Est. (Std. err.)	Est. [%] (Std. err.)	Est. [%] (Std. err.)
Science						
Voluntary-aided	0.065 (0.043)	0.111** (0.046)	0.106** (0.046)	0.064 (0.043)	0.047*** (0.006) [72.5]	0.042*** (0.006) [64.4]
Voluntary-controlled	-0.119*** (0.036)	-0.079* (0.041)	-0.089** (0.041)	-0.122*** (0.036)	0.040*** (0.005) [33.8]	0.030*** (0.005) [25.2]
Statement of SEN	-0.192*** (0.072)	-0.134* (0.071)	-0.148** (0.071)	-0.196*** (0.072)	0.058*** (0.014) [30.2]	0.044*** (0.014) [22.8]
FSM status	- -	- -	-0.194*** (0.024)	-0.110*** (0.022)		
SES measures included	yes	no	no	yes		

N=11130 children.
SEN: statement=233, no statement=10897.
School type: community=7270, voluntary-aided=1138, voluntary-controlled=2580, other=142.
Reference categories: school type/community, FSM/not claiming, SEN/no statement of SEN.
All regressions control for age, gender, ethnicity, EFL, KS1 attainment & KS2 assessment year.
SES measures: family income, mother's and partner's employment, one-parent family status, mother's and partner's education, and mother's and partner's social class.
Columns (1)-(4). Robust standard errors, corrected for clustering at the (KS2 assessment) school-level.
Columns (5)-(6). Bootstrap standard errors (1000 replications).
*, ** and *** indicate significance at the 10%, 5% and 1% levels, respectively.
Omitted variables bias [%] = (omitted variables bias/|estimated parameter in "true" model|)*100.
Imperfect proxy bias [%] = (imperfect proxy bias/|estimated parameter in "true" model|)*100.

References

Allen, R. and A. Vignoles (2006). What Should an Index of School Segregation Measure? CEE Discussion Paper.

Atkinson, A., P. Gregg, et al. (2006). The Results of 11 Plus Selection: An Investigation into Equity and Efficiency of Outcomes for Pupils in Selective LEA's. Bristol, Centre for Markets and Public Organisation (CMPO), University of Bristol.

Blanden, J. and P. Gregg (2004). "Family Income and Educational Attainment: A Review of Approaches and Evidence for the UK." Oxford Review of Economic Policy **20**(2): 245-263.

Blatchford, P., P. Bassett, et al. (2003). "Are Class Size Differences Related to Pupils' Educational Progress and Clasroom Practices? Findings from the Institute of Education Classs Size Study of Children Aged 5-7 Years." British Educational Research Journal **29**(5): 709-730.

Chevalier, A. (2004). Parental Education and Child's Education: A Natural Experiment, IZA Discussion Paper No. 1153.

Chevalier, A., C. Harmon, et al. (2005). The Impact of Parental Income and Education on the Schooling of Their Children, The Institute for Fiscal Studies Working Paper 05/05.

Croxford, L. (2000). "Is Free-Meal Entitlement a Valid Measure of School Intake Characteristics?" Educational Research and Evaluation **6**(4): 317-335.

Dearden, L., J. Ferri, et al. (2002). "The Effect of School Quality on Educational Attainment and Wages." The Review of Economics and Statistics **84**(1): 1-20.

DfES (2003). Pupil Progress by Pupil Characteristics: 2002. Statistics of Education. London, Department for Education and Skills.

DfES (2005a). Youth Cohort Study: The Activities and Experiences of 16 Years Olds: England and Wales 2004. Statistical First Release. London, Department for Education and Skills. 04/2005.

DfES (2005b). National Curriculum Assessment, GCSE and Equivalent Attainment and Post-16 Attainment by Pupil Characteristics in England 2004. Statistical First Release. London, Department for Education and Skills.

DfES (2006). PLASC Completion Notes for Primary Schools, Department for Education and Skills, mimeo.

DWP (2002a). Income Support Quarterly Statistical Enquiry May 2002. London, Department for Work and Pensions.

DWP (2002b). Jobseeker's Allowance Quarterly Statistical Enquiry May 2002. London, Department for Work and Pensions.

DWP (2004). Income Related Benefits Estimates of Take-Up in 2001/2002. London, Department for Work and Pensions.

Ermisch, J. and M. Francesconi (2000a). The Effect's of Parent's Employment on Children's Educational Attainment. Colchester, UK, Institute of Social and Economic Research, University of Essex.

Ermisch, J. and M. Francesconi (2000b). Family Structure and Children's Achievements, Institute of Social and Economic Research, University of Essex.

Ermisch, J. and M. Francesconi (2001). "Family matters: impacts of family background on educational attainments." Economica **68**(270): 137-156.

Feinstein, L. (2003). "Inequality in the Early Cognitive Development of British Children in the 1970 Cohort." Economica: 73-97.

Feinstein, L., D. Robertson, et al. (1998). Pre-School Education and Attainment in the NCDS and BCS. London, UK, Centre for Economic Performance.

Feinstein, L. and J. Symons (1999). "Attainment in secondary school." Oxford economic papers **51**(2): 300-321.

Gibson, A. and S. Asthana (2000). "Local markets and the polarization of schools in England and Wales." Transactions Of The Institute Of British Geographers **25**: 303-319.

Goldstein, H. (2003). Multlilevel Statistical Models. London, Arnold.

Goldstein, H. and P. Noden (2003). "Modelling Social Segregation." Oxford Review of Education **29**: 225-237.

Gorard, S., C. Taylor, et al. (2002). "Does school choice lead to 'spirals of decline'?" Journal Of Education Policy **17**(3): 367-384.

Gorard, S., C. Taylor, et al. (2003). Schools, Markets and Choice Policies. London, RoutledgeFalmer.

Gregg, P. and S. Machin (1998). Child Development and Success or Failure in the Youth Labour Market, Centre for Economic Performance Discussion Paper 397.

Gregg, P., E. Washbrook, et al. (2005). "The effects of a mother's return to work decision on child development in the UK." The Economic Journal **115**: F48-F80.

Hobcraft, J. (1998). Intergenerational and Life-Course Transmission of Social Exclusion: Influences of Childhood Poverty, Family Disruption, and Contact with the Police. London, Centre for Analysis of Social Exclusion CASEpaper 15.

Hutchison, D. (2003). "The Effect of Group-level Influences on Pupils' Progress in Reading." British Educational Research Journal **29**(1): 25-40.

Joshi, H. and G. Verropoulou (2000). Maternal Employment and Child Outcomes: Analysis of two Birth Cohort Studies. London, UK, The Smith Institute.

Kiernan, K. (1996). "Lone motherhood, employment and outcomes for children." International Journal of Law, Policy and the Family **10**: 233-249.

Levacic, R., A. Jenkins, et al. (2005). The Effect of School Resources On Student Attainment In English Secondary Schools, Institute of Education.

Machin, S., S. McNally, et al. (2004). "Improving pupil performance in English secondary schools: Excellence in Cities." Journal of the European Economic Association(2): 396-405.

Machin, S. and S. McNally (2004). The Literacy Hour. CEE Discussion Paper.

Mayer (1997). What money can't buy: family income and children's life chances, Harvard University Press.

Noden, P. (2000). "Rediscovering the impact of marketisation: dimensions of social segregation in England's secondary schools, 1994-1999." British Journal of Sociology of Education **21**(3): 371-390.

O'Connor, T. G., K. Thorpe, et al. (1999). "Parental Divorce and Adjustment in Adulthood: Findings from a Community Sample." Journal of Child Psychology and Psychiatry **40**(5): 777-789.

QCA (1998). Assessment and reporting arrangements: Key Stage 1 (1999), Qualifications and Curriculum Authority.

Rose, D. (1995). Official Social Classifications in the UK. Social Research Update. Guildford, University of Surrey.

Rose, D. and K. O'Reilly (1998). The ESRC review of government social classifications. London, Office for National Statistics.

Sammons, P., A. West, et al. (1997a). "Accounting for variations in pupil attainment at the end of Key Stage 1." British Educational Research Journal **23**(4): 490-511.

Sammons, P., S. Thomas, et al. (1997b). Forging Links: Effective schools and effective departments. London, Paul Chapman.

Schagen, I. and S. Schagen (2005). "Combining multilevel analysis with national value-added data sets - a case study to explore the effects of school diversity." British Educational Research Journal **31**(3): 309-328.

Shuttleworth (1995). "The relationship between social deprivation, as measured by individual free school meal eligibility, and educational attainment at GCSE in Northern Ireland: a preliminary investigation." British Educational Research Journal **21**(4): 487-504.

Storey, P. and R. Chamberlin (2001). Improving the take up of free school meals. London, Department for Education and Skills.

Strand, S. (1997). "Pupil progress during Key Stage 1: a value added analysis of school effects." British Educational Research Journal **23**(4): 471-487.

Strand, S. (1999). "Ethnic Group, Sex and Economic Disadvantage: associations with pupils' educational progress from Baseline to the end of Key Stage 1." British Educational Research Journal **25**(2): 179-202.

Strand, S. (2002). "Pupil Mobility, Attainment and Progress During Key Stage 1: a study in cautious interpretation." British Educational Research Journal **28**(1): 63-78.

Thomas, S., P. Sammons, et al. (1997). "Differential Secondary School Effectiveness: Comparing the Performance of Different Pupil Groups." British Educational Research Journal **23**(4): 451-469.

Wooldridge, J. M. (2002). Econometric Analysis of Cross Section and Panel Data. Cambridge, The MIT Press.

APPENDIX 1

The Effects of SES variables on Educational Attainment

Unless stated, the estimated "effects" are from OLS, or probit or logit models.

Family income

Ermisch and Francesconi (2001, table 2) found negative and significant effects of being in the bottom income quartile (but not of being in the second and third quartiles), relative to the top quartile, on the probability of achieving at least an A-level. In a recent review of evidence and approaches, Blanden and Gregg (2004, p262) concluded that,

Although not all our estimates are statistically significant, the consensus from our different approaches suggests that family income does affect educational outcomes.

A one-third reduction in *transitory* family income from the mean, about £140 a week (equivalent to moving from the median to around the 20th percentile in their data), reduced the probability of attaining a degree by around 4 percentage points. Effects of a similar magnitude were found for the probability of obtaining no GCSE A-C grades and the probability of post-compulsory schooling.[51] In a subsequent paper, Chevalier et al. (2005, p16-22), using trade union status to account for the endogeneity of paternal income and the raising of the school leaving age in 1974 to account for the endogeneity of parental education, found significant effects of *permanent* paternal income on the probability of post-compulsory schooling.

Parental employment

Joshi and Verropoulou (2000, p36-41) reported analyses for the children of the 1958 and 1970 birth cohorts. For children of the 1958 cohort, they found small negative and significant effects of mother's employment in first year of the child's life on reading scores when the child was aged 5-17, but not on maths scores. However, no significant effects were found of mother's employment when the child was aged 1-4, or of mother's or father's employment when the child was aged 5-17 (on reading or maths scores). For the 1970 cohort, small

[51] Their approaches included estimating the effects of changes in income, sibling fixed effects estimation and, following Mayer (1997), using post-childhood income as a proxy for permanent family characteristics.

negative and significant effects were found of mother's employment when the child was pre-school age on highest academic qualification, but not on maths or reading scores at age 10.

Ermisch and Francesconi (2000a), using retrospective information from the BHPS and sibling fixed effects estimation, found moderately large negative and significant effects of mother's full-time employment when the child was aged 0-5 on the probability of achieving at least an A-level. The effects of mother's part-time employment and father's employment at the same age were smaller and sometimes insignificant, but also negative. However, there were no significant effects of mother's and father's employment when the child was aged 6-10 and 11-15 on the probability of achieving at least an A-level.

In their analysis of the Avon birth cohort study (ALSPAC), Gregg et al. (2005, page F66) found that mother's full-time employment when the child was aged 0-18 months had small negative and significant effects on literacy when the child was aged 7, but not on their entry assessment and Key Stage 1 results. In addition, they found no significant effects of mother's part-time employment at the same age or of mother's employment when the child was aged 19-34 months.

In summary, there is some evidence of negative and significant effects of "early" maternal employment on children's educational achievement, but there appears to be no significant effects of mother's and father's employment during compulsory education.

One-parent family status

One-parent family status appears to have negative and significant effects on participation in post-compulsory schooling and educational qualifications, *unconditional* on economic circumstances (Kiernan (1996), Gregg and Machin (1998), Ermisch and Francesconi (2001a)). With sibling fixed effects estimation and without controls for economic circumstances, Ermisch and Francesconi (2000b, table 6) found that one-parent family status when the child was aged 0-5 had negative and significant effects on the probability of achieving at least an A-level. However, the effects of one-parent family status when the child was aged 6-10 or 11-15 were insignificant (but also negative).

Evidence on the effects of one-parent family status, *conditional* on economic circumstances, is mixed. Gregg and Machin (1998, tables III and VI) found no significant effects of one-

parent family status on participation in post-compulsory education, conditional on reports of financial difficulties, but negative and significant effects of 'one-parent family status but no financial difficulties' on educational attainment at 23. Dearden et al. (2002, table A3) found no significant effects of one-parent family status on highest qualifications at age 33, conditional on several measures of economic circumstances. For the same outcome, Kiernan (1996, table 3) found no significant effects of 'lone motherhood and employed mothers'. With similar samples, and similar controls for family income and wealth, while Ermisch and Francesconi (2001a, table 2) found negative and (marginally) significant effects of one-parent family status on the probability of achieving at least an A-level, Ermisch and Francesconi (2001b, table 5) found negative but (marginally) insignificant effects.

Parental education

Most studies using the 1958 and 1970 birth cohorts have found significant effects of mother's and father's education on children's educational attainment (e.g., Feinstein et al. (1998), Gregg and Machin (1998), and Dearden et al. (2002)). Similarly, Ermisch and Francesconi (2001, table 2) reported strong and significant effects of parents' educational attainments on the probability of achieving at least an A-level, in the British Household Panel Study (BHPS).

Chevalier (2004, p26-27), using the raising of the school leaving age in 1974 to address the endogeneity of parental education, found that one year of parental education increased the probability of post-compulsory schooling by 4 to 8 percentage points (when focusing on natural parents only). The effects of maternal education were larger for daughters and the effects of paternal education were only for sons. However, when accounting for the endogeneity of both paternal income and parental education, Chevalier et al. (2005, p16-22) found no significant effects of parental education.

Parental social class

There are large unconditional social class differences in educational attainment (e.g., Feinstein (2003) and DfES (2005a)).[52] In addition, most, but not all, studies have found significant conditional social class differences in educational attainment (e.g., Feinstein et al. (1998), Hobcraft (1998) and Feinstein and Symons (1999)).

[52] For example, only 33% of students with fathers in 'routine' occupations achieved 5+ GCSEs A*-C in 2004, compared to 53% of those with fathers in 'intermediate' occupations, and 77% of those with fathers in 'higher professional' occupations (DfES 2005a, table A).

APPENDIX 2

Extension to the Conceptual Framework

Extension to multiple unobserved variables

Suppose the "true" model of an outcome variable *y* is:

$$E(y \mid x_1,\ldots, x_K, q_1, q_2) = \beta_0 + \beta_1 x_1 + \ldots + \beta_K x_K + \gamma_1 q_1 + \gamma_2 q_2$$

where $\mathbf{x} = (x_1, \ldots, x_K)$ are observed explanatory variables and q_1 and q_2 are unobserved variables. Writing this in error form:

$$y = \beta_0 + \beta_1 x_1 + \ldots + \beta_K x_K + \gamma_1 q_1 + \gamma_2 q_2 + v$$

Write the linear projection of q_1 and q_2 on 1, **x** as:

$$q_1 = \delta_{01} + \delta_{11} x_1 + \ldots + \delta_{K1} x_K + r_1$$
$$q_2 = \delta_{02} + \delta_{12} x_1 + \ldots + \delta_{K2} x_K + r_2$$

The OLS regression of y on 1, **x** yields:

$$y = (\beta_0 + \gamma_1\delta_{01} + \gamma_2\delta_{02}) + (\beta_1 + \gamma_1\delta_{11} + \gamma_2\delta_{12})x_1 + \ldots + (\beta_K + \gamma_1\delta_{K1} + \gamma_2\delta_{K2})x_K + (v + \gamma r_1 + \gamma_2 r_2)$$

The omitted variable bias on x_k is $(\gamma_1\delta_{k1} + \gamma_2\delta_{k2})$.

Let z be a proxy for q_1 and q_2. Write the linear projection of q_1 and q_2 on 1, **x**, z as:

$$q_1 = \theta_{01} + \rho_{11} x_1 + \ldots + \rho_{K1} x_K + \theta_{11} z + r_1$$
$$q_2 = \theta_{02} + \rho_{12} x_1 + \ldots + \rho_{K2} x_K + \theta_{12} z + r_2$$

The OLS regression on y on 1, **x** and z gives:

$$y = (\beta_0 + \gamma_1\theta_{01} + \gamma_2\theta_{02}) + (\beta_1 + \gamma_1\rho_{11} + \gamma_2\rho_{12})x_1 + \ldots + (\beta_K + \gamma_1\rho_{K1} + \gamma_2\rho_{K2})x_K + (\gamma_1\theta_{11} + \gamma_2\theta_{12})z + (v + \gamma_1 r_1 + \gamma_2 r_2)$$

The imperfect proxy bias on x_k is $(\gamma_1\rho_{k1} + \gamma_2\rho_{k2})$.

The non-redundancy of z in the estimated "true" model

Suppose that z is not redundant in the estimated "true" model because q_2 is omitted. Write the linear projection of q_2 and on 1, **x** and q_1 as:

$$q_2 = \alpha_0 + \alpha_1 x_1 + \ldots + \alpha_K x_K + \lambda q_1 + \varepsilon$$

The OLS regression of y on 1, x and q_1 yields:

$$y = (\beta_0 + \gamma_2\alpha_0) + (\beta_1 + \gamma_2\alpha_1)x_1 + \ldots + (\beta_K + \gamma_2\alpha_K)x_K + (\gamma_1 + \gamma_2\lambda)q_1 + (v + \gamma_2\varepsilon)$$

The estimated omitted variables bias on x_k is $[(\gamma_1\delta_{k1} + \gamma_2\delta_{k2}) - \gamma_2\alpha_k]$ and the estimated imperfect proxy bias on x_k is $[(\gamma_1\rho_{k1} + \gamma_2\rho_{k2}) - \gamma_2\alpha_k]$. If the parameters have natural signs, then the consequence of omitting q_2 from the estimated "true" model, is that the estimated omitted variable and imperfect proxy biases suffer from *attenuation bias*.

Appendix Table 1

Distributions of SES variables Conditional on FSM Status

	FSM Status	
	Claiming	Not claiming
Family income/47m		
Less than £100 per week	0.39	0.05
£100-199 per week	0.39	0.15
£200-299 per week	0.16	0.29
£300-399 per week	0.05	0.24
£400 or above per week	0.01	0.27
Family employment/54m		
None	0.45	0.04
1 part-time	0.11	0.03
1 full-time or 2 part-time	0.28	0.35
1 full-time, 1 part-time	0.13	0.50
2 full-time	0.02	0.08
Mother's employment/54m		
Unemployed	0.74	0.37
Part-time	0.23	0.52
Full-time	0.03	0.10
Partner's employment/54m		
Unemployed	0.40	0.05
Part-time	0.04	0.03
Full-time	0.57	0.93
One-parent family status/54m		
One-parent	0.28	0.04
Two-parents	0.72	0.96

Appendix Table 1 (continued)

Distributions of SES variables Conditional on FSM Status

	FSM Status	
	Claiming	Not claiming
Mother's education/pregnancy		
CSE/None	0.41	0.15
Vocational	0.17	0.11
O-level	0.31	0.39
A-level	0.10	0.24
Degree	0.01	0.12
Partner's education/54m		
CSE/None	0.39	0.16
Vocational	0.13	0.10
O-level	0.23	0.25
A-level	0.22	0.31
Degree	0.03	0.19
Mother's social class/pregnancy		
Unskilled	0.06	0.02
Partly-Skilled	0.21	0.10
Skilled Manual	0.16	0.08
Skilled Non-Manual	0.41	0.45
Managerial and Technical	0.16	0.31
Professional	0.01	0.05
Partner's social class/54m		
Unskilled	0.08	0.02
Partly-Skilled	0.16	0.09
Skilled Manual	0.45	0.30
Skilled Non-Manual	0.09	0.12
Managerial and Technical	0.18	0.36
Professional	0.04	0.11

Samples described in Table 3.