International Journal of Educational Research 49 (2010) 49–68

Contents lists available at ScienceDirect

International Journal of Educational Research

journal homepage: www.elsevier.com/locate/ijedures

Pupil composition and accountability: An analysis in English primary schools

Hugh Lauder [a,*], Daphne Kounali [b], Tony Robinson [c], Harvey Goldstein [d]

[a] Education Department, University of Bath, Bath BA2 5JD, United Kingdom
[b] University of Bristol, School of Clinical Sciences, South Bristol Education Centre, Upper Maudlin St, Bristol BS2 8AE, United Kingdom
[c] Department of Mathematical Sciences, University of Bath, Bath BA2 5JD, United Kingdom
[d] Centre for Multilevel Modelling, Graduate School of Education, University of Bristol, 2 Priory Road Bristol BS8 1TX, United Kingdom

ARTICLE INFO

Article history:
Received 13 October 2008
Received in revised form 6 July 2010
Accepted 29 July 2010

Keywords:
Pupil composition
Accountability
Value added
League tables

ABSTRACT

This paper examines the role of social class and prior achievement composition effects on pupils' progress and the judgements made about school performance. It finds that there are small but significant compositional effects on pupil progress raising questions about how pupils can best be allocated to schools. Comparisons between the official contextual value added model and one that includes composition variables show that school rankings are significantly changed in the latter suggesting that official value added analyses are misleading. This study, therefore, poses a fundamental challenge to policy makers over the determinants of pupil progress and school performance and the way schools are judged, pointing to the need for a re-appraisal of policy relating to these matters.

1. Introduction

This paper raises questions about educational policy with respect to the effects of pupil composition on pupil progress and measures of accountability by studying pupil composition in a sample of English Primary schools and contextualising the findings within the educational policies of the previous New Labour Government in England. The general issues, however, are more broadly relevant to educational systems elsewhere that involve elements of public accountability through judgements made about the performance of schools and students.

There has been considerable debate about the nature and effects of pupil composition, on individual pupil and school performance, however that social class compositional effects may influence the way schools are judged has not, to our knowledge, been previously considered. Underlying this study is the question of whether the influence of composition effects is more pervasive than policy makers have been prepared to acknowledge. By composition effects we mean the effects the student body may have on school outcomes in addition to individual pupil characteristics such as their social class, gender, and ethnicity backgrounds and whether they have learning difficulties. The debate has been 'alive' since the publication of Coleman et al.'s (1966) celebrated report because it is central to two related concerns: the nature of school effectiveness and appropriate policies to raise school effectiveness. With respect to the former, Thrupp and Hursh (2006) have argued that we can identify two apparently opposed positions. The first claims that school effectiveness is a function of school management and teacher performance, while the latter claims that social factors (e.g., social class) determine pupil outcomes in schools. In this respect, pupil composition can be seen as one social factor that may be significant in determining pupil outcomes. However, they note that we can consider these two positions as at the ends of a spectrum and that much of the debate centres on the relative contributions of schools and teachers and social factors.

* Corresponding author.
E-mail address: H.Lauder@bath.ac.uk (H. Lauder).

0883-0355/$ – see front matter © 2010 Elsevier Ltd. All rights reserved.
doi:10.1016/j.ijer.2010.08.001

In policy terms, the debate is crucial because if indeed it were the case that school management and teacher performance are key factors determining school effectiveness, then the focus would be on the policies that would best raise school performance. It can be argued that policy makers have focussed, over the past 20 years, on these factors by enlisting the support of some school effectiveness and improvement studies (Goldstein, 2001). Some policy makers have even claimed that reference to social factors, is no more than an excuse for poor performance made by educators (Thrupp, 1998).

In England and to some extent the United States this has led to two specific sets of policy: (a) what may be called the state theory of learning (Lauder, Brown, Dillabough and Halsey, 2006) and (b) the introduction of market mechanisms. The state theory of learning has been a primary focus of developments in England, but also underpinned by a bipartisan government commitment to the use of market mechanisms. It is based on the idea that a combination of the repeated high stakes testing of pupils, a national curriculum with mandated pedagogy, and the publication of school rankings or 'league tables' will raise 'standards'. High stakes testing with publication of results is meant to hold schools and increasingly teachers to account while it is also intended to provide feedback for students and parents. Students and schools are set targets related to the tests and their progress is monitored in relation to them. These policies presuppose a theory of motivation in which children are stimulated to achieve high test results while teachers similarly have the spur of achieving high test results since their school will be judged against others in the published league tables.

This then raises the second issue of the role of pupil composition in public judgements about schools. Schools have usually been judged only in terms of their overall test results and more recently 'value added' measures have been accepted by policy makers since they are more closely related to pupil progress and hence it is argued to school policies. In this paper we are able to compare measures of social class, including school social class composition, family structure and income with official studies which have used only limited contextual measures of value added (CVA). In contrast, the CVA measures take into account gender, special educational needs, ethnicity, eligibility for Free School Meals, where a pupil's first language is other than English, mobility, when students move at non-standard times, age, whether pupils are in care and an Income Deprivation Affecting Children Index which measures the proportion of children under the age of 16 in an area living in low income households (Goldstein & Leckie, 2008). One reason for making this comparison is that there remain major issues as to how the official CVA measures can be used (Goldstein & Leckie, 2008). The CVA measures adopted by the DCFS are derived from data collected by the Pupil Annual School Census database (PLASC) maintained by the Department for Children, Schools and Families (DCFS) where the only compositional variable that even approximates social class composition is the proportion eligible for free schools meals.[1] In a separate paper (Kounali, Robinson, Goldstein, & Lauder, submitted for publication) we have explored the problems associated with using that variable, and especially its low sensitivity in identifying those economically deprived. We note that the inclusion of compositional or indeed any variables measured at the school level, may be relevant to judging schools for the purposes of accountability, for example when carrying out local or national inspections. For the purposes of school choice, however, it is not appropriate to include such variables since interest lies only in comparing school 'effects' and not in 'explaining' how such effects may be caused (Goldstein & Leckie, 2008). It is clear, however, that the way in which school performance is judged is important because where schools do not achieve targeted test results, a battery of measures can be externally imposed on a school in an attempt to secure better test results (Lauder, Brown, Lupton, Hempel-Jorgensen, & Castle, 2006), raising questions about teacher's professional autonomy and morale.

2. The debate

The literature on the effects of pupil composition has been extensive and while it is probably fair to say that the balance of evidence favours the existence of such effects (Teddlie & Reynolds, 2000; Palardy, 2008; Dumay & Dupriez, 2007; Leckie & Goldstein, 2009), there is no consensus (Thrupp, 1997; Nash, 2003). After three decades of studies reporting either the presence or absence of composition effects attention has turned to the basis for disagreement and these have focussed on both theoretical and methodological issues. Various explanations have been given for compositional effects. These include the following: peer group cultures; changes in the curriculum to adjust for the nature of the student composition and school policies (Thrupp, 1999). Key to these accounts is the idea that the nature of pupil composition creates a framework for responses by staff with respect to the curriculum and policy.

However, Thrupp (1999) has also noted countervailing effects in cases where schools are more socially mixed: for example, working class and/or black pupils who feel alienated within a context where there may be a predominance of middle class, mainly white, students and therefore do not 'buy in' to the dominant educationally committed ethos of the school. These countervailing effects point up why understanding the impact of student composition is so difficult. In light of these countervailing effects, which may well serve to reduce the quantitative estimates of compositional effects in some schools, Thrupp, Lauder and Robinson (2002) have argued that such effects are most likely to be observed in schools with a high proportion of pupils from the professional middle class or working class.

However, Thrupp's theoretical work arose out of the study of secondary schools and it is not immediately obvious that his theory has application in primary schools largely because while we might expect to see issues of alienation, discipline and

[1] For further details see the SCFS website: For further details see the SCFS website: http://www.standards.dfes.gov.uk/performance/1316367/CVAinPAT2005/.

social control as being significant in some schools (Hempel-Jorgensen, 2009), these are unlikely to coalesce around sub-cultures of resistance in the sense, described for example, by Willis (1977).

The contrary view as to the presence and nature of composition effects has been most consistently advanced by Nash (2003) and Nash and Harker (2006), who argues that the causes of what we observe in schools may lie outside the school and composition effects may be one example. He cites Bourdieu et al. (1999) who argues that:

> [t]he perfectly commendable wish to see things in person, close up, sometimes leads people to search for the explanatory principles of observed realities where they are not to be found (not all of them, in any case), namely, at the site of observation itself (p.181).

Nash's critique is directed at ethnographic studies such as Thrupp's and not at quantitative studies which he sees as the essential precursor to qualitative studies which seek to explain observed quantitative effects.[2]

There are several points to make in thinking about studies investigating compositional effects to emerge from this debate. Firstly, causes that can be attributed to school effects as opposed to wider societal effects are always a matter of theoretical contest, especially in relation to those processes which appear to cross the border between school and society (Lauder, Jamieson, & Wikeley, 1998). This is one reason why studies of school effectiveness should be theoretically driven. Secondly, this study focuses on the identification of composition effects rather than their causes which typically would also require qualitative analysis.[3] Finally, and most importantly for this study we need to unpack the notion of social class that is being used because it is germane to the two positions outlined above and more directly to present government policies: in particular, whether we can distinguish between four components that are often associated with social class: occupation, education, income and wealth. There has been considerable debate about the nature of social class and its operationalisation (see e.g., Devine, Savage, Scott, & Crompton, 2005; Lareau & Conley, 2008). However, social class is typically understood as relating to elements of economic capital, which can be understood in terms of market position (Goldthorpe & McKnight, 2004), income and wealth. In addition cultural capital has been included in studies because of the theoretical relationship identified by Bourdieu between culture, class and educational performance. In this study we are able to include measures which stand as proxies for some of these dimensions of class; in particular, we have data on occupation which provides some understanding of market position and income, education as a proxy for cultural capital and home ownership as a proxy for wealth.[4] While it will be apparent that these variables do not map directly on to all the dimensions of social class they do take us further than many studies that seek to identify key factors in individual pupil and school performance. In addition to measures of social class and school social class composition we shall also study the role of prior achievement at the individual level and in terms of school composition.

3. Testing the pupil composition thesis in primary schools

There are several reasons why a study of primary schools might be considered a particularly stringent test of the different elements of the pupil composition thesis. Firstly, given the view that it might well be the creation of pupil sub-cultures of resistance that are the source of a composition effect, for the reasons given above, they are more likely to be absent from primary schools. Secondly, one of the reasons why this might be the case is that primary schools tend to be small and pupils are unlikely to avert the 'gaze' of teachers. Hence, even if sub-cultures of resistance were nascent within the primary school they are less likely to develop. Thirdly, pupils in primary school may not have generated the identities necessary to create groups which challenge the teachers' and school's goals. However, at the organisational level, because primary schools are smaller schools the compositional effects on the organisation may be larger, and by the same token, the issues raised by composition may be easier to handle. In the event, there have been few large scale studies of school effectiveness in primary schools that have taken composition into account and where they have the analysis using social class has been relatively under developed (e.g. Mortimore, Sammons, Stoll, Lewis, & Ecob, 1989).

These considerations provide a theoretical framework for this study. However, in addition to the theoretical debate, there has been a related debate about methods. This latter debate is concerned with the extent to which conflicting methods and especially measurement error could give rise to dubious claims over compositional effects. It is these methodological differences, we argue, that have led to disagreement over the presence and nature of compositional effects.

4. The methodological debate

There are two major issues with respect to methodology that can explain the unresolved nature of the debate over compositional effects. These relate to the techniques and sampling used in order to identify compositional effects and which have sometimes been termed phantom effects (Harker & Tymmms, 2004). We examine the first issue by looking at the early work of Hauser (1970) who first raised some of the key issues and then the more recent paper by Harker and Tymmms (2004).

[2] Although, Thrupp (1999) is well aware of this problem noting that there may be factors that are *school based* but not *school caused* (p. 5).

[3] The qualitative work we have undertaken on a sub-sample of the schools in this study suggests that the causes may be more complex than reported by Thrupp (1999) (Lauder, Brown, Hempel-Jorgensen, & Lupton, 2008).

[4] The latter has been largely unexplored in the sociology of education but it may be that as (and when house prices rise) this gives home owning parents a stake in society which then translates into aspirations and educational hopes for their children. A discussion about the effects of wealth on education has been initiated by Conley (2001) but it is limited to its effects on a college education in the United States.

In analysing 11 schools and 900 students Hauser (1970) argued that the detection of contextual, including compositional, effects may be an artefact of the lack of well specified data at the individual level. However, his argument effectively reduces to the observation that un-modelled 'confounders' may explain some or all of the effects shown by the variable of interest. This argument, nevertheless, applies to any variable, either at the individual or school level and as such does not amount to a coherent argument against using contextual variables. He also makes the error of confusing the 'contextual effect' with the 'ecological fallacy' – in fact these are quite distinct issues. The latter refers to the 'level' at which measurements enter a statistical analysis and not which measurements are used.

Harker and Tymmms (2004) repeat the confusion of 'ecological fallacy' with 'contextual fallacy'. They employ two data sets to explore the issue of phantom effects. A data set from the Progress At School Study from New Zealand (Harker & Nash, 1996) and data from the Performance Indicators in Primary Schools Project (PIPS) (Tymms, 1999) from Britain. The New Zealand data consist of only 37 schools which can only provide limited insights, especially since as Harker (2004) subsequently acknowledged the sample of schools in his study was skewed: it did not have a representative sample of predominantly high and low social class schools. The more extensive PIPS data analysis uses the mean baseline (BLA) score as a compositional variable and finds no effect. It shows that correcting for measurement error will generally increase the compositional 'effect' although typically not by very much – depending on the number of students on which the average is based. Furthermore, it is not clear from the analysis what the true reliability of the BLA is. The figure quoted in the paper is for inter-rater reliability but the true reliability includes natural day-to-day variation, setting variation, test item variation etc. In addition, they seem to assume that it is just the mean of a predictor that matters, yet other measures such as those of spread are also important, as we shall argue. Also, interactions between compositional and other variables may be important. The paper also echoes the Hauser argument about 'confounders' but again this is a well-understood issue and not in any way specific to compositional effects. Interestingly, they fail to mention the Steedman (1980) analysis of the National Child Development Study (NCDS) data which remains one of the most powerful results in favour of compositional effects. It shows that the average prior achievement of pupils in secondary schools interacts in complex ways with individual achievement when achievement is measured at the end of secondary schooling. Steedman's findings are consistent with other studies which have shown compositional effects which can lead to important policy dilemmas when an attempt is made to translate them into practice (Thrupp, 1999). Thus, neither of these papers presents a convincing argument against either measuring, or interpreting, compositional effects as real.

With respect to sampling, following the discussion in Thrupp et al. (2002), we believe our study conforms to the following specifications:

First, the sample includes some schools from both ends of the socio-economic spectrum, this is important, as Thrupp suggests, school compositional effects are unlikely to appear in reasonably well-mixed schools because there may be countervailing factors involved.

Second, a full set of entry-level variables, including prior achievement variables, are included. Entry-level variables include a range of measures of social class, this has rarely been the case in England and Wales where a measure of eligibility for Free School Meals (FSM) typically has been used. We noted above that we have shown this measure to be unreliable in identifying disadvantaged pupils and as a predictor of subsequent performance (Kounali et al., submitted for publication; Goldstein, Kounali, & Robinson, 2008).

Third, a combination of compositional variables (e.g., prior achievement and social class composition) is used in order to measure the various dimensions of pupil composition. In addition to measures associated with social class, we have measures for prior achievement composition effects. Here the hypothesis is that it is not only factors associated with social class, such as cultural capital but also the composition of the student body in terms of prior achievement that needs to be considered since the context created by prior achievement composition may influence pupils' orientation to learning.

Fourth, different techniques for measuring composition should be used. The typical measure employed is the mean, for example, of the social class composition. Additionally, ratios of high to low social class proportions could be used or other measures of 'spread'. In this study we use ratios because they better capture the distribution of pupils within schools. For example, from our qualitative study of a subsample of 12 of the schools in this sample, it is clear that some school compositions comprise a significant proportion of pupils from professional backgrounds and working class backgrounds. These tend to be on the rural fringes of the urban centre which is the focus of this study (see below). Given these considerations we move to a description of the sample.

5. Study design: the HARPS project

The HARPS project is an acronym for 'Hampshire Research with Primary Schools' and looks at the impact of school composition upon student academic progress. The main aim of the study is to estimate and better understand *compositional effects* at the primary school level. Compositional variables included in this study will be; social class (Appendix 2),[5] prior achievement, class organisation with respect to age-mix, school registration rates and school turbulence measured by the proportion of new pupils arriving and leaving school at non-standard times.

[5] Social class has been classified according to the Goldthorpe and Hope (1974) scale, this operationalises a theory of class, rather than socio-economic status, however as an abbreviation we have used the term SES in the following tables.

The research design is both quantitative and qualitative. The project design has 3 nested parts:

1. A large scale analysis of over 300 primary schools.
2. A study of a subsample of 46 schools in one urban school area.
3. Detailed case studies of 12 schools.

In the present study we focus on the subsample of schools for which detailed socio-economic data were collected through the HARPS project. In particular, we perform value added analysis on pupil's attainment using the Qualifications and Curriculum Authority tests for Year 3 (grade 4) in reading and mental arithmetic[6] (QCA3 tests) in this subsample of schools whilst taking into account the history of attainment of these pupils from the beginning of school and socio-economic status at year as well as the standard characteristics typically employed by the DCFS.

For this subsample of 46 schools of the 300 primary schools in Hampshire, we have collected and analysed detailed family background information from the year 3 children in these schools. The subsample contains family background data on 1653 year 3 pupils from a total of 2012 students attending 46 out of all 50 schools in an urban area, which we have called Greentown, during the second semester of the academic year 2004–2005. Relevant to economic status these data include: occupational group (Goldthorpe & Hope, 1974), working status; home ownership; whether in receipt of Working Tax Credit; whether in receipt of FSM, level of education of the parents, and house movements during the child's lifetime.[7]

In Britain pupils are tested at approximately age 5 with respect to numeracy and literacy, this is known as the Baseline test and then in English and Maths at age 7 (designated the end of 'Key Stage 1': these tests are referred to as KS1 tests); and English, Maths and Science at ages 11 and 14 (end of 'Key Stage 2' and 'Key Stage 3' referred to respectively as KS2 and KS3 tests). Descriptive statistics on the prior attainment of these pupils at both baseline and at Key Stage 1 (KS1) according to SES can be found in Appendix 1 (Tables A1 and A2). In Appendix 1 we also include detailed description of the pupils attainment at QCA3 in reading and mental arithmetic in Fig. A1. Descriptive statistics of all the characteristics measured and included in the value added analysis are detailed in Tables A3 and A4. These include characteristics of the children (age, gender, special education needs' provision (SEN), English as second language (EAL), absences, home changes and school changes and class vertical groupings' individual membership: vertical groupings are when older and younger pupils are taught together) as well as parental characteristics (economic indicators: renting their home, in receipt of working tax-credit, social class measured by the occupational classification (Goldthorpe & Hope, 1974, Appendix 2), level of education, single parenthood and family size, Table A3).

The deprivation geography of Hampshire according to the multiple deprivation index (Noble et al., 2004) suggests that the children attending the selected Greentown schools are not among the most deprived in Hampshire, but this area nevertheless includes pockets of particularly deprived areas, thus in principle, covering the deprivation spectrum.

The majority of the responders were the mothers of the children (87%) or the mother and the father (1.3%) and <1% was not the child's parent. Female responders accounted for 90% of the returned questionnaires. This is also a sample that is predominantly white with 92.7% of the responders being white-British or Irish, another 3.4% being white-mixed and another 3.3% all other ethnic backgrounds.

In this paper we have used value added analyses of QCA 3 tests in reading and mental arithmetic accounting for prior attainment at both baseline and KS1 tests. Reading competence, it can be argued, is essential for all forms of education although it is also related to social class and cultural capital (Nash & Harker, 2006). In contrast, learning in maths is typically seen to be more subject to the influence of the school and less of the home.

6. Statistical methods – modelling achievement and progress

We have fitted a multilevel model with the QCA reading and mental arithmetic test scores as joint responses and the set of variables as predictors given in Tables 1–3.

A formal statement of the model is as follows:

$$
\begin{aligned}
y_{ij}^{(1)} &= \beta_0^{(1)} + \beta_1^{(1)} x_{1,ij}^{(1)} + \sum_{k=2}^{p} \beta_k^{(1)} x_{k,ij} + u_{0j}^{(1)} + u_{1j}^{(1)} x_{1,ij} + e_{ij}^{(1)} \\
y_{ij}^{(2)} &= \beta_0^{(2)} + \beta_1^{(2)} x_{1,ij}^{(2)} + \sum_{k=2}^{p} \beta_k^{(2)} x_{k,ij} + u_{0j}^{(2)} + u_{1j}^{(2)} x_{1,ij} + e_{ij}^{(2)} \\
&\begin{pmatrix} u_{0j}^{(1)} \\ u_{1j}^{(1)} \\ u_{0j}^{(2)} \\ u_{1j}^{(2)} \end{pmatrix} \sim MVN(0, \Omega_u) \quad \begin{pmatrix} e_{ij}^{(1)} \\ e_{ij}^{(2)} \end{pmatrix} \sim MVN(0, \Omega_e)
\end{aligned}
\tag{1}
$$

[6] This is the body in England charged with the management of the national curriculum and student assessment.
[7] Details of how these data were collected are in Brown and Thrupp (2005).

Table 1
The effect of school composition on the attainment at QCA3 tests in reading and mental arithmetic.

Predictor compositional characteristics	Scale-coding	QCA3 reading score (scale: 0–36 points)			QCA3 mental arithmetic score (scale:0–15 points)		
		Mean	SE		Mean	SE	
Registration rates	3-year school average of (number of pupils on Roll) over (school capacity) Scale: 0.6–1.3 Modal value 0.9	4.56	1.74	**	2.00	1.14	*
% of new-pupils	Proportion of new pupils (% pupil changing schools at a different postcode i.e. excluding changes to aligned Infant–Junior schools) Scale: 0–1, mean value = 0.16 SD = 0.29, modal value = 0.06	−1.23	0.96		0.84	0.60	
SES school differentials	Number of students with low SES over Number of students with high SES 25% centile = 1.2, mode = 2.5 75% centile = 6.2	−0.08	0.02	***	0.002	0.02	
School baseline score average	Baselineschool average score Normalized (normal equivalent deviate)	0.57	0.27	***	−0.40	0.18	**
School KS1 attainment differentials	KS1 attainment school Differentials (proportion level w/1) Over (proportion level 3); Scale [−2 2]	0.44	0.24	*	0.37	0.17	**
Interaction of type of Maths-test administered and KS1 attainment class differentials	KS1 attainment class differentials for those pupils who took the more difficult test 2B	0.27	0.33		−0.44	0.15	**

* Statistical significance level 5%.
** Statistical significance level 2.5%.
*** Statistical significance level <1%.

where i indexes individual pupils and j indexes schools and the superscript indexes whether reading or mathematics. There are p fixed predictors (x) in this model of which the first is the KS1 score, which also has a random coefficient at the school level and is the baseline reading or maths score as appropriate. Model (1) is fitted as a bivariate model. This is efficient since it fully utilises all the data, including where a pupil only has one QCA3 score (3.8%). The results are presented separately for maths and reading scores.

For the variable 'special needs' we carried out preliminary analyses relating the categories to the QCA3 test scores. There are only small mean differences among the special needs groups so that we have used a binary split 'special needs/no special needs' for this predictor. We have carried out similar analyses for social class and find an approximate linear relationship with simple category scores 1, ..., 4. Likewise for main carer's education a linear relationship for maths requires scores 1,2,2,3 and for reading scores 1,2,3,6. Thus for these latter variables we have a single covariate which are these category scores. For the year grouping variable we have formed a composite variable: no mixed ages at either year 2 or year 3; not mixed year 2 and mixed year 3; not mixed year 3 and mixed year 2; mixed both years. Additional exploration of year grouping effects is the subject of further research.

7. Results

In this section we discuss the factors that we found to have an effect on QCA3 test scores of reading and mental arithmetic. These include both individually measured characteristics as well as compositional variables at school and class-level. The effects of all individual pupil characteristics along with those of their parents are detailed in the Appendix 1, Tables A5 and A6, respectively. These tables explore these effects, fitting model (1), where each table presents a subset of the variables fitted.

It is worth noting the strong effect of all socio-economic indicators on reading. This is not, however, the case for mathematics performance. Interestingly vertical groupings effects were found. These are difficult to interpret and are discussed separately in Appendix 3.

Table 2
Statistics for the estimates of the between school variation in attainment on QCA3 tests as well as residual (unexplained) variance and correlations.

Random parameters	Mean (SE) correlation (cor.)	Mean (SE) correlation (cor.)	Mean (SE) correlation (cor.)	Mean (SE) correlation (cor.)
	Mean	SE		
School level between school variation	KS1 reading prior attainment below level 3	KS1 mental arithmetic prior attainment below level 3	KS1 reading prior attainment at level 3	KS1 mental arithmetic prior attainment at level 3
KS1 reading prior attainment below level 3	2.12 (**) (0.76) cor. = 1.0			
KS1 mental arithmetic prior attainment below level 3	0.13 (0.38) cor. = 0.002	1.39 (0.38) cor. = 1.0		
KS1 reading prior attainment at level 3	−2.16 (***) (0.84) cor. = −0.82	−1.10 (**) (0.50) cor. = −0.57	2.54 (**) (1.10) cor. = 1.0	
KS1 mental arithmetic prior attainment at level 3	0.01 (0.32) cor. = −0.02	−0.94 (***) (0.30) cor. = −0.94	0.69 (*) (0.41) cor. = −0.71	0.59 (0.26) cor. = 1.0
Residual variation – observation level				
Attainment in reading variance				17.23 (***) (0.75) cor. = 1.0
Covariance between attained reading and mental arithmetic				1.45 (***) (0.26) cor. = 0.17
Attainment in mental arithmetic variance				4.11 (***) (0.18) cor. = 1.0

* Statistical significance level 5%.
** Statistical significance level 2.5%.
*** Statistical significance level <1%.

We shall focus on the compositional effects found and which are detailed in Table 1. We should note here that prior attainment composition at baseline was defined as the school average at baseline. The prior attainment composition at KS1 was defined differently. This is because KS1 test results are measured on a discrete scale and are highly skewed towards higher levels with 63% of pupils attaining level 2A and above. As such a mean does not provide a good measure of KS1 school composition. This is the reason why we choose the ratio of those attained the lowest level over those attaining the highest KS1 level. This is proportional to the proportion of low performing pupils in a school with reference to the high performing ones and a measure of balance/mix or differentials of prior KS1 attainment. We used the same approach for defining the school socio-economic composition. We should also note that these measures for most schools represent class rather than school differentials as most schools were too small to have more than one year 3 class.

We also look in detail at the extent of between-school variation in Table 2.

8. Social class composition effects

After adjusting for the other individual and compositional factors, social class composition as measured by the Goldthorpe and Hope (1974) occupational scale, was found to have a significant impact on pupils' reading attainment only. Children going to schools with high proportions of pupils from working class and/or non-'paid work backgrounds perform significantly worse in the reading test. The social class composition is defined as the ratio of the number of working class pupils to the number of those from professional backgrounds. The point gap between pupils attending schools at the 25% lowest centile and the 75% upper centile of the school distribution of this social class ratio is almost half a point that is 0.05 standard deviations. No effect was found for mental arithmetic. This may be because social background may have little or no effect, since in contrast to literacy maths, including mental arithmetic is taught in schools: the differences in compositional effect can be seen in terms of the transference of cultural capital in which reading is central, whereas maths may not be. Support for this interpretation may also be seen in the performance of pupils for whom English is a second language (Table A5). While pupils for whom English is a second language perform less well in reading they perform better in mental arithmetic.

9. Prior achievement composition effects

With respect to prior achievement composition effects we need to make two observations. Firstly, prior achievement cannot be wholly divorced from social class composition for reading because the influence of social class while clearly strong at the point when the baseline tests were administered (Table A1), most likely remains, although as we noted above there will be differences in effects between reading and mental arithmetic. Secondly with respect to these data we need to take a developmental perspective. Pupils from working class and families not in paid work backgrounds start their school careers well behind those from professional social class backgrounds as shown in Table A1. Those lower achievers who attend high baseline intake classes perform better in reading and maths than those who do not. There is no advantage for higher math achievers in the same class. This result is consistent with Steedman's (1980) findings and suggests that composition does matter at a particularly crucial stage in a child's education.

However, for those judged to be of higher ability in mathematics (those who take the more demanding exam at KS3[8]), there is a penalty in being mixed with those judged of lower prior achievement. This was also found by Roberston and Symons (2003). In Fig. 1 we compare composition effects for students from professional and working class and those not in paid work backgrounds with different prior achievement profiles who have achieved level 3 at KS1 and those below level 3 in the context of schools with different forms of social class and prior achievement compositions.[9]

It can be seen from Fig. 1 that with respect to reading low achieving, low social class pupils do better in a high attaining high SES school than in a low achieving low SES school. Equally high attaining high SES pupils in a low attaining low SES school do marginally worse than if in a high attaining high SES school which raises questions about the implications for policy. The pattern of predicted differences for mental arithmetic are quite similar with one important exception. We see that pupils in schools with high prior attainment intakes (that is both baseline and KS1 tests) perform worse than pupils in schools with low attainment intakes. This compositional effect in maths is clear among low attaining pupils. This is because predicted differences are based on school composition on attainment at the difficult test for the high attaining pupils because

[8] Two types of mathematics tests were administered. Each of these batteries of tests included mental arithmetic as a common component. It is for this reason that in this analysis we only considered this component of mathematics test. The two tests were designed for different levels of mathematics ability with Test 2A being the easiest. 15 of the participating schools opted for the more difficult test while 2 of these schools took both.

[9] Working class pupils were included who are renting and pupils from professional families who own their homes. High prior attainers are those in the 75% centile of the baseline score distribution and level 3 at KS1 and low prior attainers are those in the 25% centile of the baseline distribution and level 1 or below at KS1. For the school compositional prior achievement variables, those at the 1% and 99% centiles were included. A high prior attaining school had both high (above the 99th centile) baseline average, low proportion of KS1 level 1 attainers and high registration rates with average for turbulent pupils. A low prior attaining school has low baseline average, high proportion of level 1 attainers and low registration rates.

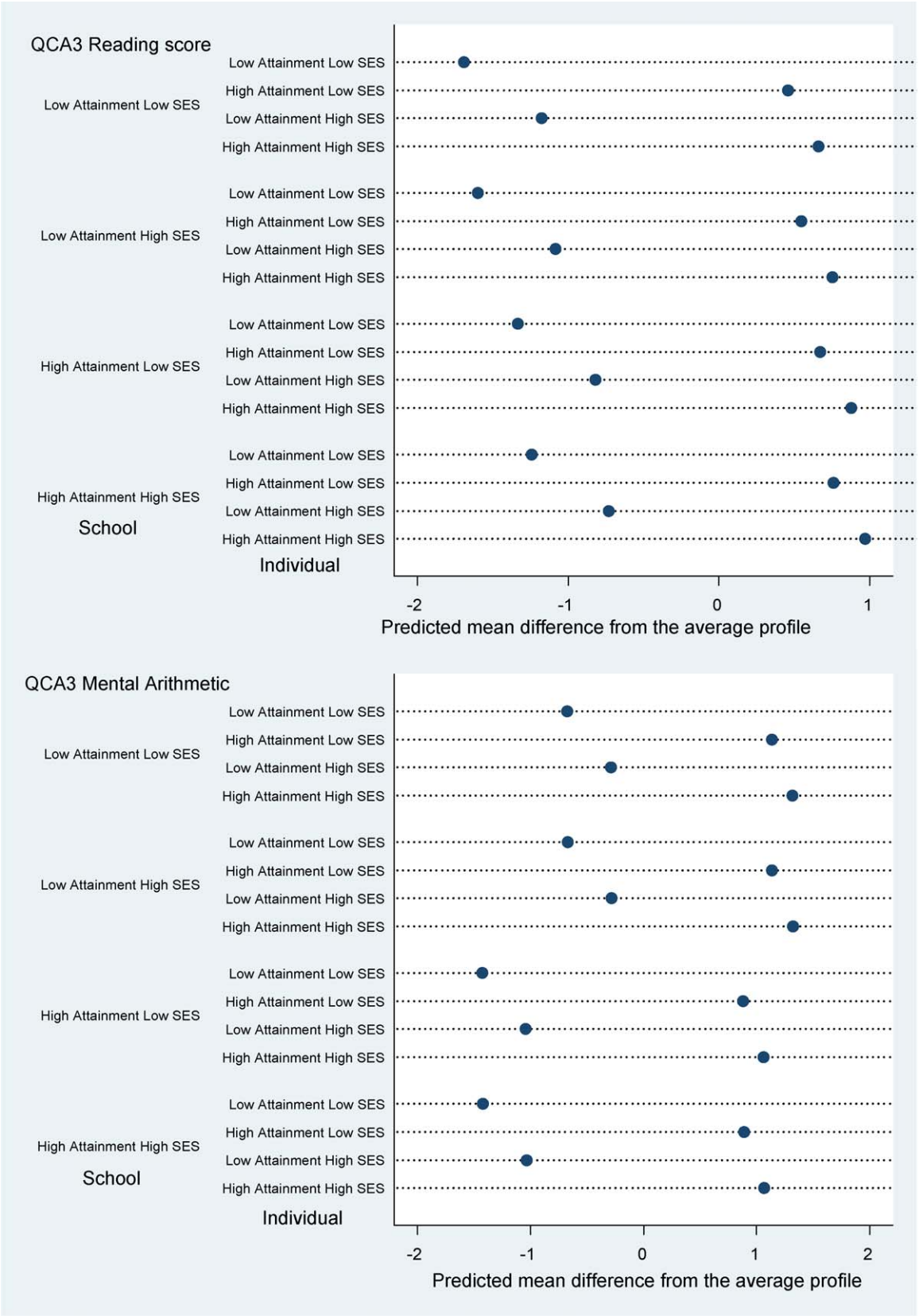

Fig. 1. Expected difference from the average pupil-profile in predicted reading scores according to different pupil profiles.

most of high attaining school opted for the difficult test. Thus, this correctly reflects the compositional effect of prior ability in mathematics, described above.

The finding, in principle, may open a window of optimism in relation to cohort studies such as Feinstein's (2003) which suggests that by the time low achieving low SES pupils start school there is no improvement in their relative position. The effect here is small and clearly requires further research. However, the policy implications of this finding are complex. If a policy of increasing the number of low achievers in a class is followed, then the mean baseline score of the class is reduced with the implication that low achievers will do less well. Broadly speaking there are two ideal-type policy options: (a) dividing pupils between high achieving and low achieving classes, in effect setting and (b) having mixed ability classes. Neither option provides a straightforward 'win win' possibility, hence, the need to work out the predicted scores of pupils in schools with different compositions.

Finally, the data on composition effects are almost certainly underestimates for two reasons. Our missing data appear to be biased towards those on low incomes (Goldstein et al., 2008) so our analysis is likely to reduce the composition effect sizes. We argued earlier that it is in schools at the extremes of the social class distribution that composition effects are most likely to be observed, however, our sample may not fulfil this criterion adequately. In our sample, while there were schools with such compositions they were relatively few in number. Six schools had over 40% of pupils from professional backgrounds, two of which had 60%. In contrast, approximately 11 schools had predominantly working class populations. Of these only four schools had a predominance of those who are unemployed or low skilled routine workers.

A further hypothesis, worthy of consideration, concerns the way that composition is reconfigured by grouping strategies (Roberston & Symons, 2003). They found powerful grouping effects related to social class, although they did not place these effects within the wider context of a multi-level model. It may well be that composition effects are therefore mediated by grouping strategies. In our study there were effects which were suggestive of the influence of grouping strategies which we refer to below. In Lauder et al. (2008) an explanation, drawing on the qualitative data from this study, is given of these composition effects in the light of schools' setting and grouping strategies (see also Appendix 3).

10. School effectiveness and accountability

The second issue to be investigated in this study is that of the influence of composition measures on school rankings when compared to the standard contextual value added measures used in official statistics. We noted earlier that there is a view amongst some policy makers that any reference to social factors such as school composition is merely an excuse for poor school performance (Thrupp, 1998). When we look at the measures by which schools are judged this view appears inscribed in traditional league tables unadjusted for social factors. However, while contextually value added league tables at least acknowledge the influence of some social factors, the question is whether they do sufficient justice to the challenges schools confront with respect to composition variables.

Here there are two issues worthy of note. The first is whether the more detailed measures of social class, including social class composition, that we have been able to deploy produce different results to official league tables. The second concerns the assumption made in both traditional and contextually value added league tables that schools add value uniformly among all pupils. Once we add our measures of social class and prior achievement composition, do schools add value uniformly for all pupils or are some schools better at teaching some groups of pupil than others? This is a significant question once it is acknowledged that the social composition of a school can influence pupil achievement, especially in schools where pupils are drawn from heterogeneous backgrounds with respect to social class and prior achievement.

In what follows we present data patterns on school performance according to our model, which includes compositional variables, and compare it to the standard model typically used by the UK Department for Schools, Children and Families (DCFS). We are particularly interested in two aspects of this comparison. Firstly, with respect to school performance for students who have achieved level 3 at KS1 and those who have achieved below this level.[10] As indicated, the policy assumption is that schools can be good schools for all pupils and if they are not, they should be. But league table rankings do not take into account the possibility that some schools are 'good' for some pupils and perhaps not others. In addition, the term 'good' is a relative one and it is perfectly possible for all schools to be 'good' even though they can still be ranked. Secondly, when we compare the school rankings according to our model and the DCFS contextual value added model it will be clear that school rankings are changed. This same point is made by Leckie and Goldstein (2009). When these two sets of data patterns are compared important questions are raised about the reliability of the basis for official judgements about school performance.

In Figs. 2 and 3 we compare the DCFS contextually value added league table with our model which includes social class and prior achievement composition variables for reading and mental arithmetic. In our model we distinguish between value added measures for reading and mental arithmetic for pupils below level 3 at KS1 and those who have attained level 3 at KS1 in order to see whether there are differential effects for different types of pupil.

[10] We have allowed for differential school effectiveness by including a random coefficient for the (dummy) variable associated with the highest level category at KS1. This allows us to obtain two 'value added' effects for each school: one that applies to the pupils who are in the highest scoring category at KS1 and one for the remainder.

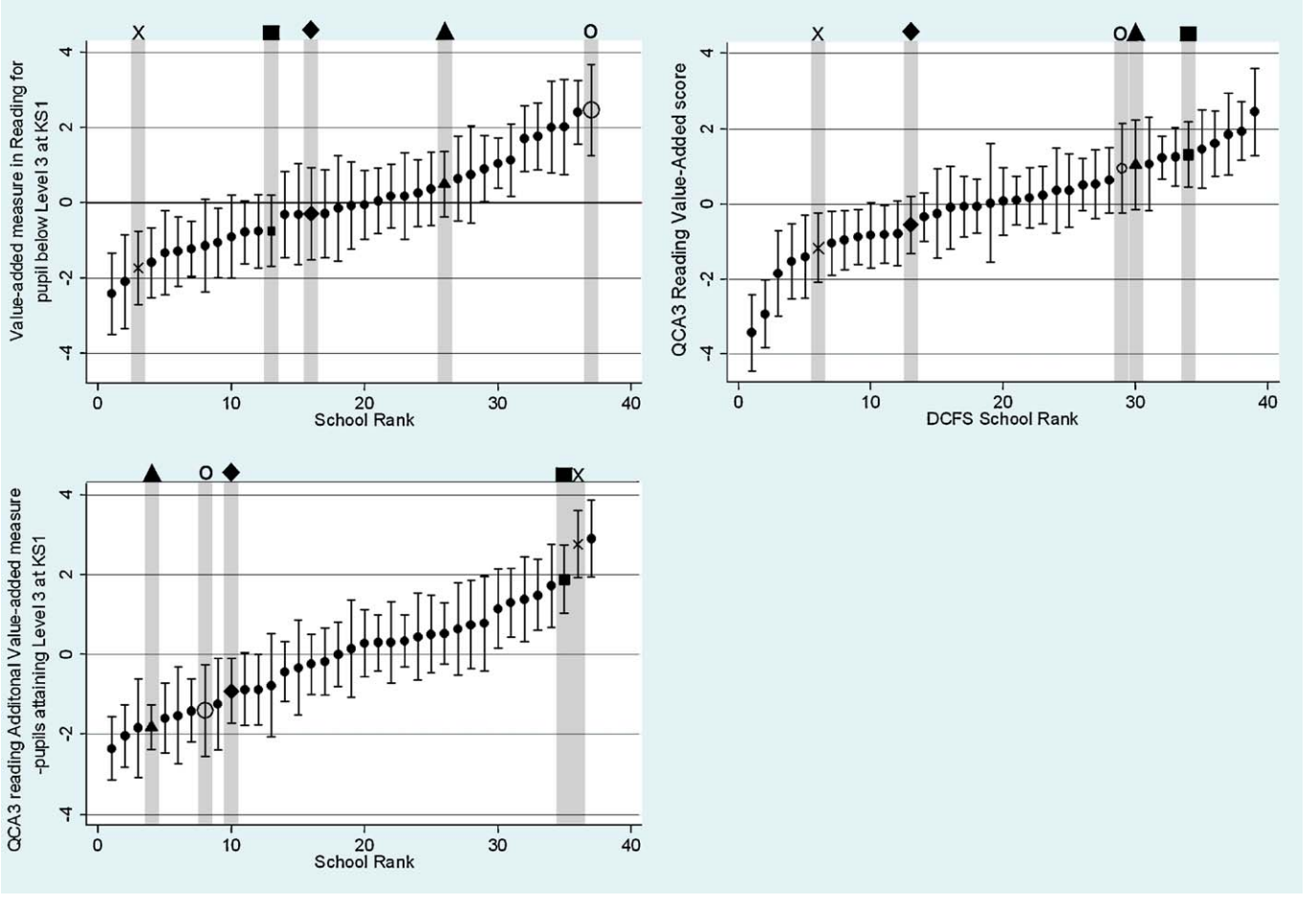

Fig. 2. (a) School differential effectiveness in reading for low and high achievers taking into account compositional variables. (b) School differential effectiveness in mental arithmetic for low and high achievers taking into account compositional variables.

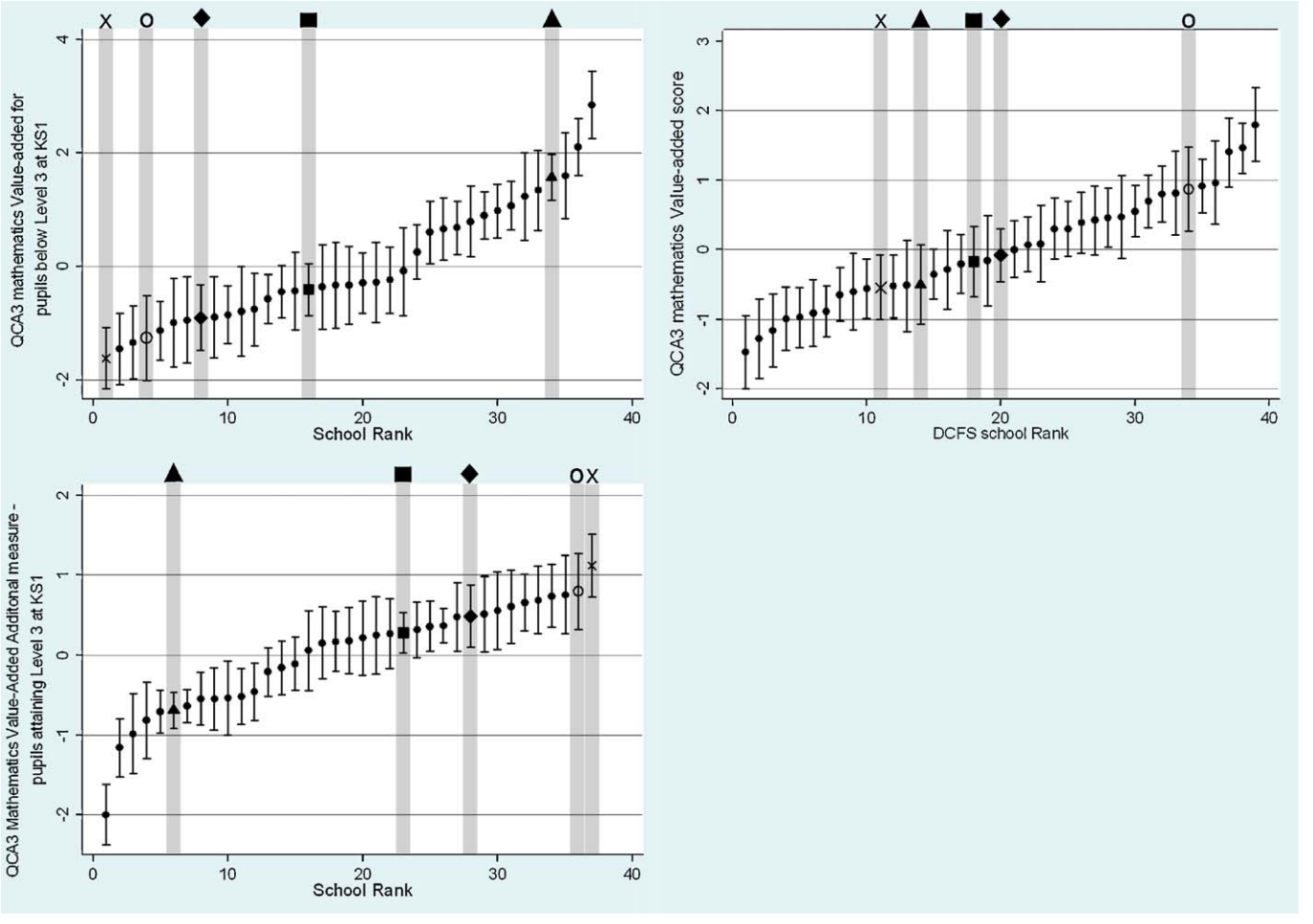

Fig. 3. (a) and (b) Detailed comparison in value added scores between two analyses for QCA3 reading and mental arithmetic: one which considers detailed socio-economic and compositional data and the typical analysis undertaken by DCFS.

When we compare our rankings with those of the DCFS (Figs. 2 and 3) we can see that as regards reading the DCFS ranking are quite different to ours. Consider school X. On the DCFS ranking it is low in the rankings, as it is on our analysis for low achievers but on our analysis the same school does very well for high achievers.

A similar point can be made for this school in maths. It seems as if it does very well with high achieving pupils but less well with low achieving pupils. Inspection of Figs. 2 and 3 shows that school X is not alone in performing quite differently for high and low achievers when compared to the DCFS ranking. When we consider the school denoted by the triangle, it is quite high in the DCFS rankings for reading and lower for mental arithmetic. However in our model it has a somewhat lower position for pupils who are below level 3 for reading but performs less well for those pupils who had attained KS1 level in reading. In contrast for mental arithmetic it is performs less well according to the DCFS league table, while in our model it does very well for the below level 3 pupils, although it is towards the bottom for those that have attained KS1 level 3.

Four points arise from this analysis. Firstly, schools that perform well on reading do not necessarily perform well in mental arithmetic and vice-versa. Secondly, schools do perform differentially according to the pupils they are teaching. Thirdly, when the DCFS contextual value added ranking are compared to our analysis then the rankings change when our additional measure of social class composition is taken into account and for the differential effects that schools have on particular groups of students particularly in terms of prior achievement. Finally, all value added estimates will have large confidence intervals attached to them that severely restricts their usefulness. In particular, if these value added estimates are to be used for school choice it is the predicted values several years into the future that are appropriate, not current school values and as (Goldstein & Leckie, 2008) point out these have much higher confidence intervals.

Cast in these terms the differences between schools are not surprising. Although, that value added scores are fairer indicators than raw averages because they compare equals or "more equals" (these adjust the raw attainment means for differences between schools in intake and other factors taken into account in their estimation), their usefulness is limited. School rankings based on the value added scores are still severely constrained by very large uncertainties that usually make most schools indistinguishable from one another in test performance (Goldstein & Spiegelhalter, 1996).

Two points of interpretation of these comparisons emerge from these findings. The first is methodological. Our composition effects are small, yet they make a significant difference to school rankings, this attests to the sensitivity of measures used to construct league tables as well as reflecting the large confidence intervals referred to above. Secondly, there is a substantive issue concerning the observation that different schools have differential effects on different groups of pupils. Can this be explained by compositional effects or are they more likely, for example, to be teacher effects?

Ideally we should be able to follow Raudenbush and Willms (1995) in distinguishing between Type A and B effects. The former concerns contextual factors, including pupil composition, while the latter refers to the effects of school practice, including leadership, curriculum, content and instructional processes. However our data set does not include quantitative measures relating to Type B effects which may have enabled better judgements to be made about these different explanations. However, it is quite likely that it would be difficult to distinguish between these two explanations since it may well be that better teachers are attracted to schools with higher social class intakes and that better teaching practices are enabled in schools with higher social class intakes. Qualitative evidence for the latter hypothesis is provided by Hempel-Jorgensen (2009) in her study of four of the 46 Hampshire schools in our sample. That said, it will be clear from the above figures that working class schools and their teachers appear to do very well for some groups of pupils. Clearly this is an area for further investigation.

11. Conclusion

This study of social class and prior achievement composition effects raises fundamental questions about education policy with respect to how pupils are distributed between schools according to social class and prior achievement and with how schools are judged. In both cases they pose difficult dilemmas for policy makers.

While the composition effects that we have identified are small, there are good reasons for thinking they are underestimates. Our findings suggest that value judgements are fundamental to the question of how best to allocate pupils between schools. This is because our results show that, for example in reading, gains for one group, in this case low achievers from working class backgrounds, will occur alongside reductions for high achieving pupils from the professional middle class. By the same token the latter do relatively better in schools with similar achieving pupils. In principle, policy makers could make the decision either to separate high performing middle class pupils from lower achieving working class pupils, or they could opt for a system of mixing according to prior achievement, as the Inner London Education Authority attempted to do with its system of banding according to ability (West, 1997).

However, in practice policy makers have attempted to arrogate this issue to parents through the problematic notion of parental choice, via geographical mobility or selection among locally available schools. Allocation of students is, therefore, no longer a matter of decision making by politicians but a function of parental choice and the various forms of market power they can command. Once pupils have been allocated to schools through the market, then the state often intervenes by emphasising the importance of various kinds of grouping practices, often leading to the redistribution of pupils according to social class and attainment within schools. Clearly, parental choice will only be relevant to student outcomes if, as Thrupp and Hursh (2006) have noted, it can be assumed that the determinants of school outcomes lie with school management systems and the quality of teaching, rather than with school composition factors.

This brings us to our second issue. The question of school performance is not only about the limits and possibilities set by the allocation of pupils to schools but also about how schools are judged. The quality of predictors used in models that

compare schools is important for value added analysis. No set of indicators can fully account for the strengths and weaknesses of schools and in this obvious sense they are proxies for school processes and outcomes. However, the question is whether they are fair or misleading proxies: if the latter then some schools and teachers may suffer a considerable injustice.

In comparing our measures of school ranking, including compositional variables, with those used in English official contextual value added measures it is quite clear that the inclusion of compositional variables changes the rank order of schools. In turn this prompts fundamental questions about the fairness in judging schools according to official measures. When this finding is coupled with questions about the reliability of FSM measures (Kounali, Robinson, Goldstein, 2008) then we can be rightly sceptical about key elements in accountability policies operated by recent governments.

In addition, it is also clear that we should be wary of the concept so popular in policy and media circles of a 'good' school. What this study shows is that schools may be 'good' with some groups of children and not with others but these differences are masked by the one-dimensional rank ordering of schools in league tables.

This study reiterates previous concerns that pose a fundamental challenge to policy makers over the determinants of pupil progress and school performance and the way schools are judged, pointing to the need for a re-appraisal of policy relating to these matters.

Acknowledgements

We would like to thank the Editor of the IJER and the referees for their insightful comments. This paper could not have been written without the critical comments and insights of our colleagues on the project Ruth Lupton, Martin Thrupp, Ceri Brown, Amelia Hempel-Jorgensen and Frances Castle. We should acknowledge the role of Ruth Lupton in providing critical insights through the writing of this paper and Ceri Brown in the development and administration of the questionnaire on family background and structure which was crucial to this paper. This study could not have been undertaken without the close cooperation of Hampshire LA and in particular Nigel Hill and Paula Guy. We would also like to thank Dick Harker who has acted as a critical friend.

Appendix 1. Sample descriptives

See Fig. A1, Tables A1–A6.

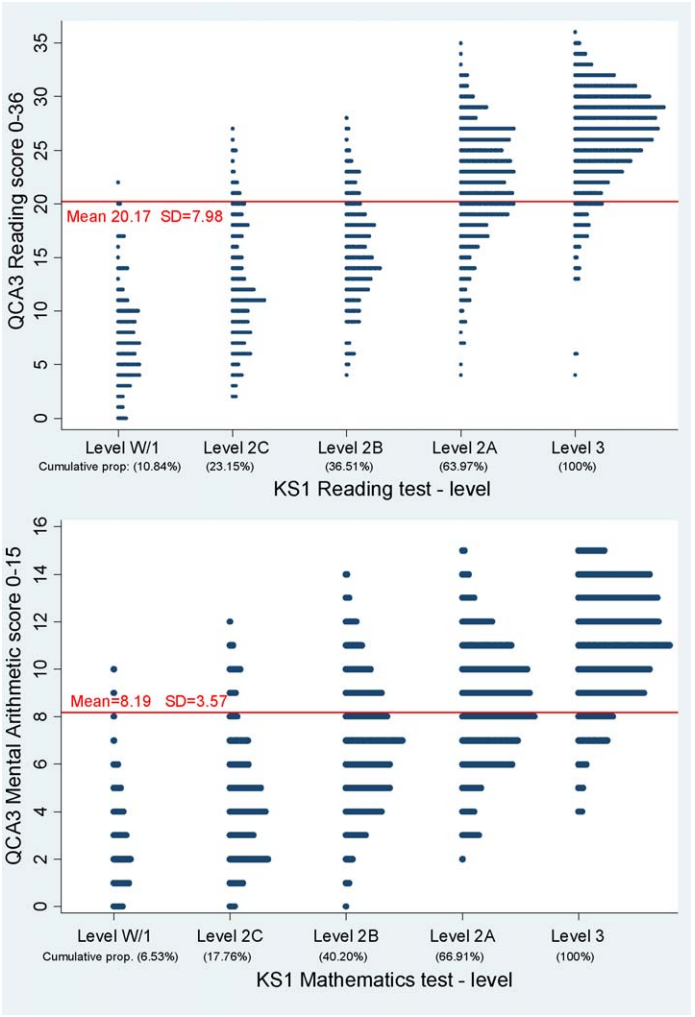

Fig. A1. QCA3 test results in reading and mental arithmetic.

Table A1
Mean differences (in standard deviations) relative to children from non-working families (FSM recipients) in baseline literacy scores according to social class.

SES (occupational class)	Literacy (standard deviation units)		Numeracy (standard deviation units)	
	Mean (SD)	95% confidence interval	Mean (SD)	95% confidence interval
High	1.06 (0.11)	[0.84 1.27]	1.06 (0.11)	[0.85 1.28]
Middle	0.78 (0.09)	[0.59 0.97]	0.81 (0.09)	[0.62 0.99]
Low	0.40 (0.09)	[0.21 0.59]	0.45 (0.09)	[0.27 0.64]

Table A2
Proportion of pupils attaining level 3 at KS1 tests of literacy and numeracy according to social class.

SES (occupational class)	KS1 Literacy	KS1 Numeracy
	% attaining level 3	% attaining level 3
High	53%	48%
Middle	45%	42%
Low	30%	27%
Not working	13%	15%

Table A3
Descriptives of the characteristics of children and their parents in the studied sample.

Variable	Units	Mean (SD)	Range
Child characteristics			
Age	Years	4.5 (0.29)	[3.9 5.9]
Absences	(half-day sessions missed – log-scale + 1)	2.3 (1.02) 50% of children did not have any absences	[0 5.2]

Variable	Units	*N* (%)
Child characteristics		
School changes	Binary	374 (19%)
Home changes	Binary	542 (28%)
SEN	Binary	457 (24%)
EAL	Binary	44 (2.3%)
Male	Binary	961 (49.5%)
Vertical groupings		
No age mix	Binary	1133 (65.2%)
Only at year 3	Binary	213 (12.3%)
Only at year 2	Binary	111 (6.4%)
Age-mixed both years	Binary	280 (15.1%)
Parental characteristics		
Renting	Binary	425 (22%)
In receipt of working tax credit	Binary	479 (29.2%)
SES		
High	Binary	247 (15.6%)
Middle	Binary	587 (37%)
Low	Binary	620 (39%)
Not working	Binary	131 (8.3%)
Highest level of education attained		
Secondary <16 years	Binary	455 (28%)
Secondary 16–19	Binary	228 (14%)
Further/vocational	Binary	643 (39%)
University graduate/postgraduate	Binary	313 (19%)
Single parent	Binary	280 (17.1%)
Family size		
1 child	Binary	157 (9.6%)
2 children	Binary	778 (47.5%)
3 children	Binary	416 (25.3%)
>3 children	Binary	240 (17.7%)

Table A4
Number of families renting their home and receiving working tax credit according to SES.

SES (occupational class)	Not renting		Renting		Total	
	Not in receipt of working tax credit	In receipt of working tax credit	Not in receipt of working tax credit	In receipt of working tax credit	Not in receipt of working tax credit	In receipt of working tax credit
High	187	43	11	6	198	49
Middle	403	107	50	27	453	134
Low	265	154	87	114	352	268
Non-working	23	1	97	10	120	11

Table A5
The effect of demographic, prior achievement and class organization characteristics on the attainment at QCA3 tests in reading and mental arithmetic.

Predictor	Scale-coding	QCA3 reading score (scale: 0–36 points, mean = 8.19, SD = 7.98)			QCA3 mental arithmetic score (scale: 0–15 points, mean = 8.19, SD = 3.57)		
Fixed part of the model		Mean	SE		Mean	SE	
Demographics							
Age at baseline	Standardized (i.e. mean = 0, SD = 1)	0.31	0.14	**	−0.08	0.07	
Special education needs provision (PLASC 2005)	0: no special provision; 1: school provision + statement	−2.95	1.83		−1.72	1.01	*
English as second language	0: no; 1: yes	−2.39	0.90	**	0.53	0.43	
No of half-day sessions absent	log-transformed	0.11	0.13		0.04	0.07	
Gender	0: female; 1: male	−0.33	0.27		0.96	0.13	***

Table A5 (*Continued*)

Predictor	Scale-coding	QCA3 reading score (scale: 0–36 points, mean = 8.19, SD = 7.98)			QCA3 mental arithmetic score (scale: 0–15 points, mean = 8.19, SD = 3.57)		
Fixed part of the model		Mean	SE		Mean	SE	
Prior attainment							
KS1 level 2C	Reference category KS1 level W/1	−0.04	2.42		0.92	1.51	
KS1 level 2B		2.86	2.38		2.90	1.52	*
KS1 level 2A		6.60	2.28	**	2.41	1.51	
KS1 level 3		10.77	2.32	***	4.27	1.53	***
Type of maths test taken	Type 2A: reference; type 2B; both[a]	1.56	0.34	***	1.83	0.19	***
Both.qca3r		1.46	1.23		2.07	0.62	***
Baseline score (Hampshire standard)	Standardized (i.e. mean = 0. SD = 1)	0.55	0.79		−0.45	0.51	
Interaction baseline score with KS1 attainment level 2C		1.56	1.02		1.19	0.58	**
Interaction baseline score with KS1 attainment level 2B		0.31	0.90		1.24	0.54	**
Interaction baseline score with KS1 attainment level 2A		0.54	0.84		1.12	0.53	**
Interaction baseline score with KS1 attainment level 3		−0.10	0.81		1.25	0.53	**
Vertical groupings (VG: measured at individual level)							
Taught in mixed class only during year 3	Reference category is a non-age-mixed class at both years	0.90	0.53	*	0.87	0.36	**
Taught in mixed class only during year 2		0.87	0.69		1.33	0.41	***
Taught in mixed class only during both year 2 and year 3		−1.95	0.51	***	0.43	0.32	

* Statistical significance level 5%.
** Statistical significance level 2.5%.
*** Statistical significance level <1%.
[a] Two types of mathematics tests were administered. Each of these batteries of tests included mental arithmetic as a common component. It is for this reason that in this analysis we only considered this component of mathematics test. The two tests were designed for different levels of mathematics ability with Test 2A being the easiest. !5 of the participating schools opted for the more difficult test while 2 of these schools took both.

Table A6
The effect of parental background and mobility factors on the attainment at QCA3 tests in reading and mental arithmetic.

Predictor parental background	Scale-coding	QCA3 reading score (scale: 0–36 points)			QCA3 Mental arithmetic score (scale:0–15 points)		
		Mean	SE		Mean	SE	
Renting their home	0: no; 1: yes (among those not in receipt of working tax credit)	−2.49	1.30	*	0.10	0.63	
Social class (SES based on occupational classification)	1: high; 2: middle; 3: low; 4: not working (among those who are not renting)	−1.55	0.80	*	−0.50	0.57	
Level of education	1: second <16; [a]2: second 16–19; 3: further/vocational; 6: University graduates + postgraduate	0.30	0.09	***	−0.21	0.11	*
Single parent	0: no; 1: yes	−0.31	0.38		−0.16	0.19	
Family size	1: 1 child; 2: 2 children; 3: 3 children; 4: more than 4 children	−0.55	0.15	***	0.03	0.07	
Interaction of renting by receipt of working tax credit	Effect of renting one's home among those in receipt of working tax credit	1.38	0.65	**	0.11	0.32	
Interaction of renting with SES	SES effect (modification) among those who are renting	0.79	0.43	*	−0.01	0.21	
Interaction of SEN with SES	SES effect (modification) among those with SEN	0.63	0.56		0.17	0.25	
Interaction of SEN with education	Parental education effect (modification) among those with SEN	−0.37	0.27		0.01	0.30	

Table A6 (*Continued*)

Predictor parental background	Scale-coding	QCA3 reading score (scale: 0–36 points)			QCA3 Mental arithmetic score (scale:0–15 points)	
		Mean	SE		Mean	SE
Interaction: SES * (difference) level w/1 – level 2C	SES effect (modification) according to KS1 attainment	1.47	0.86	*	0.42	0.57
Interaction: SES * (difference) level w/1 – level 2B		0.94	0.83		0.15	0.58
Interaction: SES * (difference) level w/1 – level 2A		1.60	0.80	**	0.37	0.58
Interaction: SES * (difference) Level w/1 – level 3		1.21	0.81		−0.16	0.19
Mobility indicators (at individual level)						
School changes	New-pupil to a school (i.e. changed schools at a different postcode) 0: no change; 1: school change	−0.11	0.13		−0.03	0.06
Home changes	Changed home 0: never; 1: once; 2: twice; … 5: more than 5 times	0.71	0.69		−0.34	0.33

* Statistical significance level 5%.
** Statistical significance level 2.5%.
*** Statistical significance level <1%.
[a] The coding for the parental level of education for mathematics is different as follows: 1: second. <16; 2 second. 16–19; 2: further/Vocational.; 3: University graduates + postgraduate

Appendix 2

In this section, we provide some details on the classification system used to characterize social class, having recorded occupation categories using the Goldthorpe occupation-scale (Goldthorpe and Hope, 1974).

Social class	Questionnaire category
High	Professionals
Middle	Managers/administrators; associate professionals
Low	Skilled craftsmen; clerical/secretarial; sales; machine operatives; personal and protective services
Not working	From employment data recording lack of work for both of the carers

Note: The occupation of both carers at present and in the past was recorded and used for assessing social class as follows:

We calculated family social class as the current occupation of the male carer and the current occupation of the female carer in the absence of a response from the male carer. We compared different methods of combining current and historical occupational information from both carers. The best method (less biased or inconsistent with the official statistics) was found be the one based on the occupation of the father. Based on the data collected, we outline below the factors which could lead to such biases i.e. when the highest occupational class is used among carers at present or historically.

a. Adopting the widely used strategy of considering the highest occupational class between carers resulted in exaggerated representation of the professional and managerial occupational groups when compared with data with the Hampshire and national statistics on occupation – with the associated proportions almost twice as high as those reported in the county-wide national statistics.
b. Also we found that almost 45% of the occupation codes determining the family's social class (as the highest occupation in the couple) were those of the male responders or partners. It is also interesting to note, that in the occupational classes associated with the highest and middle SES (as defined in Table 1) the proportion of male-determined codes was close to the average while the lowest and missing or unemployed classes were predominantly determined by females. In those later low SES classes a significant proportion (45% of clerical/secretarial; 49% of sales/machine operatives/personal and protective services) and 67% of the non-responders and unemployed) were single parents. It is clear that family structure (i.e. single parenthood) is associated with social class where the proportion of single parents in the higher social class occupations is 7%, compared to 11.3% and 26% in the middle and lower social class occupations, respectively.
c. Also, we found that the majority of responses on the highest occupational category refer to the past (64.4%). We also see that the majority of the current ones (55.7%) refer to the occupation of the male bread-winner from high occupational categories and the majority of past ones (61.1%) refer to female bread-winner from low occupational categories. This suggests that the bread-winner is male. If we look closer at the change of occupational status for the major bread-winner we find that those with higher social class occupations suffer less in the job market (job-stability/insecurity). A total of 365 families (22.1%) experienced a worsening of their occupational status. Among these families, 81% corresponds to female

bread-winners. Among higher social class occupations 20.7% experienced a worsening of their occupational status compared with 23.7% and 24.3% for the middle and low social class occupations. The gender of the bread-winner modifies this relationship and suggests that working mothers might experience a tougher deal in the job market. More specifically, we find that if we control for the gender of the major bread-winner then among females with occupations associated with high social class 27.4% experience worsening of their occupational status. This worsening of occupational status is 36.9% and 39.2% among women with middle and low social class occupations, respectively.

Appendix 3

Data were available for individual pupils at both years 2 and 3 indicating whether the pupil was taught in a class where pupils from older, younger or both older and younger year groups were included. This gave us data patterns related to those vertical groupings which are often employed in primary schools where schools are small or there are falling rolls and/or scarce resources and in which class size needs to be balanced with cost. Grouping strategies in primary schools are complex, varied and subject to change (Lupton et al., 2006) and therefore it was only in this limited sense that we were able to see if there are any grouping effects.

The picture we have identified is complex (Table A5). Mixed age classes at both year 2 and 3 appear to have a significant negative effect on QCA3 reading and no impact on the mental arithmetic score (pupils in mixed age classes in both years scored at QCA3 reading 0.24 standard deviations lower compared to those taught at classes with no age mixing). Age mixing at year 2 was found to have a positive effect on both reading and mental arithmetic (pupils in mixed age classes at year 2 had an advantage of 0.1 and 0.24 standard deviations on QCA3 test results for reading and mental arithmetic respectively against those taught at classes with no age mixing). Age mixing during the year of QCA3 tests was also found to have a positive effect for mathematics only (pupils in mixed age classes at year 3 had an of advantage 0.4 standard deviations against those taught at classes with no age mixing).

For year 3 pupils who were in mixed classes, we do not know the extent to which teachers exercise judgement on pupils' abilities in order to assign pupils into vertical groupings when the need arises. In schools where the proportion of low KS1 attainers is <10% only 1% of pupils are mixed with younger year groups whereas in schools where the proportion of low attainers is higher (more than 10%) significantly more of their pupils (19%) are grouped with younger year groups. Similarly, at KS1, the average baseline score in both reading and maths was almost one standard deviation above the average for those pupils who were taught in classes with older year groups.

Thus the above effects relating to vertical groupings could reflect differences in performance associated with the matching of pupils' ability to a class of similar ability, i.e. selection effects. If, for example, teachers group lower ability pupils with younger year groups then this could explain the benefit these pupils experience in QCA3 attainment. It should be noted that the majority (80%) of pupils attending mixed-age classes at year 2, consist of pupils actually mixing with older year groups. Thus, the positive impact on attainment that pupils taught in such classes experienced could also reflect pupils' ability.

References

Bourdieu, P., Accardo, A., Balzas, G., Beaud, S., Bonvin, F., Bourdieu, E., et al. (1999). *The weight of the world: Social suffering in contemporary society*. Standford, CA: Stanford University Press.

Brown, C., & Thrupp, M. (2005). Pulling out all the stops: Getting a high response rate with a questionnaire for parents. *Paper presented to the 2005 BERA conference*.

Coleman, J., Campbell, E., Hobson, C., McPartland, J., Mood, A., Weinfeld, F., et al. (1966). *Equality of educational opportunity*. Washington, DC: Government Printing Office.

Conley, D. (2001). Capital for college: Parental assets and post secondary schooling. *Sociology of Education, 74*(1), 59–72.

Devine, F., Savage, M., Scott, J., & Crompton, R. (2005). *Rethinking class, culture, identities and lifestyle*. Basingstoke: Palgrave.

Dumay, X., & Dupriez, V. (2007). Accounting for Class effect Using the TIMSS Eighth-Grade Database: Net effect of Group Composition net effect of class process and joint effect. *School Effectiveness and School Imprrovement, 18*(4), 383–408.

Feinstein, L. (2003). Inequality in the early cognitive development of British children. *Economica, 70*, 73–97.

Goldstein, H. (2001). Using pupil performance data for judging schools and teachers: Scope and limitations. *British Educational Research Journal, 27*, 433–442.

Goldstein, H. (2008). Evidence and educational policy -some reflections and allegations. *Cambridge Journal of Education, 38*(3), 393–400.

Goldstein, H., Kounali, D., & Robinson, A. (2008). Modelling measurement errors and category misclassifications in multilevel models. *Statistical Modelling, 8*(3), 243–261.

Goldthorpe, J., & McKnight, A. (2004). The Economic Basis of Social Class. In S. Morgan, D. Grusky, & G. Fields (Eds.), *Mobility and Inequality: frontiers of Research from Sociology and Economics*. Stanford: Stanford University Press.

Goldstein, H., & Leckie, G. (2008). School league tables: What can they really tell us? *Significance 5(67–69)*.

Goldstein, H., & Spiegelhalter, D. J. (1996). League tables and their limitations:statistical issues in comparisons of institutional performance (with discussion). *Journal of the Royal Statistical Society (Series A), 159*, 385–443.

Goldthorpe, J. H., & Hope, K. (1974). *The social grading of occupations: A new approach and scale*. Oxford: Clarendon Press – Oxford.

Harker, R. (2004). *Compositional effects in school effectiveness studies: A New Zealand case study*. San Diego: AERA.

Harker, R., & Nash, R. (1996). Academic Outcomes and School Effectiveness: type A and B effects. *New Zealand Journal of Educational Studies, 31*, 143–170.

Harker, R., & Tymmms, P. (2004). The effect of student composition on school outcomes. *School Effectiveness and Improvement, 15*(2), 177–199.

Hauser, R. M. (1970). Context and consex: A cautionary tale. *American Journal of Sociology, 75*(4), 645–664.

Hempel-Jorgensen, A. (2009). The construction of the 'ideal pupil' and pupils' perceptions of 'misbehaviour' and discipline: Contrasting experiences in a low and a high socio-economic primary school. *British Journal of Sociology of Education, 30*(4), 435–448.

Kounali, D., Robinson, A., Goldstein, H., & Lauder, H. (submitted for publication). The probity of free school meals as a proxy measure for disadvantage, University of Bath, Maths/Education.

Lareau, A., & Conley, D. (Eds.). (2008). *Social class: How does it work?* New York: Russell Sage Foundation.
Lauder, H., Brown, P., Dillabough, J.-A., & Halsey, A. H. (2006). Introduction: The prospects for education: Individualization, globalization and social change. In H. Lauder, P. Brown, J.-A. Dillabough, & A. H. Halsey (Eds.), *Education, globalization and social change* (pp. 1–70). Oxford: Oxford University Press.
Lauder, H., Brown, C., Hempel-Jorgensen, A., & Lupton, R. (2008). Explaining composition effects in primary schools. Bath University, Education Department. http://www.bath.ac.uk/research/harps/.
Lauder, H., Brown, C., Lupton, R., Castle, F., & Hempel-Jorgensen, A. (2006). Politics and professionalism: The question of school and teacher autonomy in relation to grouping practices. *BERA Symposium on the HARPS Project*.
Lauder, H., Jamieson, I., & Wikeley, F. (1998). Models of effective schools: Limits and capabilities. In R.. Slee, G.. Weiner, & *School effectiveness for whom? Challenges to the school effectiveness and school improvement movements*. London: Falmer.
Leckie, G., & Goldstein, H. (2009). The limitations of using school league tables to inform school choice. *Journal of the Royal Statistical Society (Series A) 172 Part 4.*
Lupton, R, Hempel-Jorgensen,A, Castle, F, Brown,C and Lauder, H (2006) School socio-economic composition and pupil grouping in primary school, British Education Research Association Conference, University of Warwick.
Mortimore, P., Sammons, P., Stoll, L., Lewis, D., & Ecob, R. (1989). *School matters*. Wells, Open Books.
Nash, R. (2003). Is the school composition effect real? A discussion with evidence from the UK PISA data. *School Effectiveness and Improvement, 14*(4), 441–457.
Nash, R., & Harker, R. (2006). Signals of success: Decoding the sociological meaning of associations between childhood abilities and adult educational achievement. In H. Lauder, P. Brown, J.-A. Dillabough, & A. H. Halsey (Eds.), *Education, globalization and social change*. Oxford: Oxford University Press.
Noble, M., Wright, G., Dibben, C., Smith, G., Mclennan, D., Anttila, C., Barnes, H., Mokhtar, C., Noble, S., Avenell, D., Gardner, J., Govizzi, I., & Lloyd, M. (2004). Indices of deprivation 2004. Report for Deputy Prime Minister Office: Neighbourhood Renewal Unit, London.
Palardy, G. (2008). Differential school effects among low middle and high social class composition schools: a multiple group,multilevel latent growth curve analysis. *School Effectiveness and School Improvement, 19*(1), 21–49.
Raudenbush, S., & Willms, D. (1995). The estimation of school effects. *Journal of Educational and Behavioral Statistics, 20*(4), 307–335.
Roberston, D., & Symons, J. (2003). Do peer groups matter: School effects and academic achievement'. *Economica, 277*, 31–54.
Teddlie, C., & Reynolds, D. (Eds.). (2000). *The International Handbook of School Effectiveness*. London: Falmer Press.
Steedman, J. (1980). *Progress in secondary schools: Findings from the national child development study*. London: National Children's Bureau.
Thrupp, M. (1997). The school mix effect: The history of an enduring problem in educational research, policy and practice. *British Journal of Sociology of Education, 16*(2), 183–203.
Thrupp, M. (1998). Exploring the politics of blame: School inspection and its contestation in New Zealand and England. *Comparative Education, 34*(2), 195–209.
Thrupp, M. (1999). *Schools making a difference: Let's be realistic*. Buckingham: Open University Press.
Thrupp, M., & Hursh, D. (2006). The limits of managerialist school reform: The case of target setting in England and the USA. In H. Lauder, P. Brown, J.-A. Dillabough, & A. H. Halsey (Eds.), *Education, globalization and social change*. Oxford: Oxford University Press.
Thrupp, M., Lauder, H., & Robinson, T. (2002). School composition and peer effects. *International Journal of Educational Research, 37*, 483–504.
Tymms, P. (1999). *Baseline Assessment and Monitoring in Primary Schools*. London: Peter Fulton.
West, A. (1997). The way to a fairer admissions system. *Parliamentary Brief, 5*(1), 17.
Willis, P. (1977). *Learning to labour*. Farnborough: Saxon House.