

## **Explaining Differences in Key Stage 2 Pupil Attainments: a multilevel analysis**

PAUL MARCHANT & KATHY HALL

*Leeds Metropolitan University, UK*

**ABSTRACT** This paper examines the effects of key characteristics on achievement at the end of Key Stage 2 with reference to the results of all KS2 pupils in one LEA in the north of England. It offers new evidence of the relative impact of various pupil characteristics on performance by using generalised regression methods, which take account of the structure of the data. The paper explores simultaneously the impact on achievement of school attended and school cluster (based on groups of feeder schools) and four pupil factors; namely, gender, age, social disadvantage (defined in terms of entitlement to free school meals) and having English as an additional language. It discusses the findings with respect to other recent studies in the field.

### **Introduction and Background**

The background of this paper, and of our continuing work in this area, is the current educational policies which (a) require schools to assess pupils at various stages in their schooling, and (b) require schools and LEAs to collect and use attainment data to set targets for further attainment. The existence of the resultant rich datasets offers the potential for schools and LEAs to monitor achievement according to a wide range of variables, e.g. to monitor its own attainment over time, to compare attainment with the national picture, to compare the relative achievement of groups such as boys and girls, to single out 'underachieving' or highly successful settings for further, perhaps qualitative, exploration. To maximise the potential for such work, data has to be recorded as accurately and as comprehensively as possible. One large LEA in the north of England, with which we are currently working, is seeking to develop its systems of data recording and analyses to furnish schools and other agencies, including OFSTED, with the best quality data on which to make judgements about pupil and school performance and improvement. This paper represents the first outcome phase of a research and development collaboration with colleagues in that LEA. It is based on all primary schools with Key Stage 2 (KS2) pupils—a total of 107 schools—and on the results of standard tasks and tests in English, Mathematics and Science for 5,269 pupils in 1999.

This paper examines the effects of key characteristics on achievement at the end of KS2. The aim of the work overall is to furnish this LEA with the kind of analyses that would support it in identifying specific groups of students or schools that might need more focused attention, resources and closer monitoring to improve their performance. Since the

introduction of the National Curriculum and its assessment framework, several studies have sought to understand the influences on pupil assessment results (e.g., Schagen & Sainsbury, 1996; Sammons & Hind, 1997; Strand, 1999; Yang et al., 2000; Demie, 2001; Gorard et al., 2001; Levacic & Woods, 2002; Strand, 2002). The rationale for the current work stems from the needs of regional or local users for information that pertains to their particular context and which may not be available through national or other regional analyses, although the analysis and findings have research relevance beyond the local context of the LEA. Moreover, this paper does offer new evidence of the relative impact of various pupil characteristics on performance by using generalised regression methods, which take account of the structure of the data—pupils grouped within schools, for instance. Many studies attempting to account for pupil achievement tend not to explore key characteristics simultaneously and, as Strand (1999) notes, even when some do, interactions between characteristics are rarely made explicit. Aitkin and Zuzovsky (1994) demonstrated the danger of not incorporating interaction effects.

This paper explores simultaneously the impact of school attended and school cluster (based on groups of feeder schools) and four pupil factors: gender, age, social disadvantage (defined in terms of entitlement to free school meals) and having English as an additional language. No school-level information was given.

### **Procedures and Statistical Methods**

Schools are expected to submit two kinds of KS2 pupil assessment data to the LEA—the results of the externally-set standard assessment tasks and tests (SATs) in English, Maths and Science and the results of Teacher Assessments (TA). TA results, which are based on evidence-informed judgements of teachers, are not used as a basis for target setting and as a result tend to have far less status in practice than they had in the mid nineties (Hall & Harding, 2001) when they were typically published alongside SATs results in league tables and in school prospectuses. TA results are not part of our analyses here.

The SATs results take the form of levels of attainment which can range from 1 through 8. The range of levels within which the great majority of KS 2 pupils are expected to work is 2 to 5 and the expected attainment for the majority of pupils at the end of KS 2 is level 4. Pupils at the end of KS2 are expected to achieve at level 4 in all assessed areas. Detailed descriptions of these levels are in National Curriculum and assessment policy documents (e.g., QCA & DfEE, 1999).

In the case of Maths the level awarded is based on the results of two written tests and a mental arithmetic test. In Science the level allocated is determined on the basis of two written tests. In the case of English, pupils take three different tests—one each in reading comprehension, writing, and spelling and handwriting. A composite level based on these assessments yields an overall level for English. In their recording of results in 1999 this LEA recorded the composite level per child for English.

On receiving the data from the LEA our first task was to check the codings used and the completeness of the data set. Associated difficulties and omissions were highlighted and fed back to the LEA, specifically problems in relation to the specification (or inaccurate or confusing specification) of ethnic background, home/first language and religion. Such difficulties arose because different people had been involved in the inputting of data on the LEA database, and had, on occasions, used their own idiosyncratic methods making consistency and interpretation a problem. Moreover, schools had not always adhered to the recommended procedures in how they recorded evidence and dealt with missing values, and they had not had sufficiently specific guidance on how to record background information about pupils. This means that some records in the original dataset could not be

included in our analysis. The test results had typically 8% missing and the EAL variable was 15% unknown. In collaboration with LEA advisors, we have now drawn up a coding system for future logging of information which has already rendered more recent data sets much more accurate and complete.

It was agreed that information pertaining to variations in pupil attainment at the end of KS2 would be relevant to the LEA. More specifically, it was agreed to explore the influence of pupil background characteristics and school attended on attainment in English, Maths and Science at the end of KS2. To do this we applied multilevel modelling (MLM) which is a method that is able to take account of the inbuilt hierarchies that inevitably exist in social science data. MLM allows us to determine whether various combinations of explanatory variables add usefully to the prediction of outcomes, children’s success, and give confidence intervals around the magnitude of the variables’ effects. In our case, MLM recognises that our pupils are different ages, are both male and female, that some receive free school meals (FSM) and others have English as an additional language. We are interested in how these influence children’s success. It recognises that the children are in particular schools, which in turn are in particular clusters or pyramids of schools. It treats school and school cluster as random as opposed to fixed effects. A fixed effect model for every school, for example, would not be able to give reliable estimates of the effects because of proliferation of the number of parameters. Not to take account of the ‘nested’ nature of our cases or, put another way, their multilevel structure, would run the risk of misrepresentation and the exaggeration of significance (Paterson & Goldstein, 1991; Jones & Duncan, 1998). Most of the remainder of the paper identifies and discusses a number of significant relationships in pupils’ attainment.

**Overview of Achievement within the LEA**

Table 1 summarises the level of achievement in the three tested areas of the National Curriculum—Maths, Science and English—while in tables 2–5 these achievements are broken down by our key variables. The first thing one notices here is that the vast majority

Table 1 Overall achievement in KS2 tests

	Maths		Science		English	
	N	%	N	%	N	%
Level 3 or below	1288	27	1002	20	1295	27
Level 4 or above	3550	73	3970	80	3529	73
Total	4838	100	4972	100	4824	100

Table 2 Performance at KS2 by gender

	Maths		Science		English	
Gender	Boys	Girls	Boys	Girls	Boys	Girls
Level 3 or below	26%	27%	21%	20%	31%	22%
Level 4 or above	74%	73%	79%	80%	69%	78%
Total	2443	2395	2519	2453	2397	2427

Table 3 Performance at KS2 by age (at the end of school year)

	Maths		Science		English	
	< 11.5	> 11.5	< 11.5	> 11.5	< 11.5	> 11.5
Age	< 11.5	> 11.5	< 11.5	> 11.5	< 11.5	> 11.5
Level 3 or below	29%	24%	23%	18%	30%	24%
Level 4 or above	71%	76%	77%	82%	70%	76%
Total	2427	2411	2516	2456	2435	2389

Table 4 Performance at KS2 by FSM status

	Maths		Science		English	
	FSM	nonFSM	FSM	non FSM	FSM	nonFSM
Level 3 or below	39%	23%	31%	17%	40%	23%
Level 4 or above	61%	77%	69%	83%	60%	77%
Total	1288	3550	1002	3970	1295	3529

Table 5 Performance at KS2 by EAL status

	Maths		Science		English	
	EAL	nonEAL	EAL	non	EAL	nonEAL
Level 3 or below	39%	26%	41%	18%	39%	27%
Level 4 or above	61%	74%	59%	82%	61%	73%
Total	1146	2936	890	3314	1166	2903

of pupils in this LEA achieve at the expected level for their stage and age, i.e., they obtain at least a level 4 in each assessed area. It is noteworthy that this government has set a national target of 80% of pupils obtaining at least level 4 in English and 75% of pupils obtaining at least a level 4 in Maths by 2002. This 1999 cohort in this LEA does not currently meet this target. No national targets are specified for Science, although 80% of our cohort obtains at least level 4 in Science. Of concern for the LEA is that between one-fifth and just over one-quarter of pupils do not achieve the expected level in the tests.

A closer investigation of the effects of key characteristics on achievement, using MLM, provides an explanation for variations in this achievement profile. As already noted, one of the advantages of this approach is that it allows a simultaneous examination of the data, including the effect of interaction i.e., the joint effect of two or more predictors being not just simply the sum of the individual effects.

We also handle the three subjects in a multivariate way, because every child has the opportunity to take the tests in the three subjects. That is: they are the same children taking all three tests, and not a different sample of children taking each of the tests. This allows the correlations of children's performance on the tests to be calculated.

**Assessing the Impact of Key Characteristics**

These four pupil characteristics—age (at the end of the school year), gender, whether in receipt of free school meals (FSM), and whether the pupil has English as an additional language (EAL) or, more specifically, whether English is not the pupil’s home language—have varying effects on achievement. In our analysis age is a continuous variable, whereas the other three are binary. We sought to answer the question: what is the probability of a pupil obtaining at least L4 in each subject, given her or his characteristics of age, gender, FSM, and EAL status?

In examining the effects, we took as our base line or reference point an eleven-year old boy, not on FSM and without EAL status. For example, in focusing on the impact of age (of being one year older) on English achievement, we get  $2 \times 2 \times 2 \times 2 = 16$  combinations. This gives 48 combinations overall for the three subjects. Table 6 presents the results. We should point out that the numbers (more strictly the coefficients) in table 6 are not probabilities themselves, but are related. They are in fact contributions to the logarithm of the odds. The odds of achieving L4 is given by the probability achieving L4 (p) divided by the probability of not achieving it, (1-p), i.e.  $p/(1-p)$ . It is exactly the equivalent to the odds used horse-racing. The logarithm of odds is potentially unbounded, unlike probability, bounded at 0 and 1, or odds, bounded at 0. The logarithm of the odds,  $\log((p/(1-p)))$ , is known as the logit. There are technical reasons for working with logits rather than probabilities although we can convert from one to the other as follows:  $\text{probability} = p = \exp(\text{logit}) / (1 + \exp(\text{logit})) = (1 + \exp(-\text{logit}))^{-1}$ . This is used below.

The coefficient values in table 6 are point estimates of logits and these are subject to uncertainty. The standard errors are typically 0.2.

As already stated, the reference category is an 11-year-old boy not on FSM and without EAL. The logits for obtaining at least level 4 in English, Maths and Science are respectively 0.571, 0.861 and 1.561, for such a pupil. We can take the anti-logit to give the three probabilities of passing. These are: English 0.639, Maths 0.709, Science 0.826. These probabilities are different from the proportions shown in table 1 as these latter figures are

Table 6 The coefficients of the model found

---

The intercept terms for the reference category are the logits for an 11-year-old boy, not on FSM or EAL:  
 English:0.571, Maths: 0.861 and Science: 1.561

---

		Age	Female	Free school meal	English as an additional language
Age	English	<b>0.433</b>			
	Maths	<b>0.632</b>			
	Science	<b>0.614</b>			
Female	English	Not inc	<b>0.693</b>		
	Maths	Not inc	-0.078 NS		
	Science	Not inc	0.052 NS		
FSM	English	<b>0.608</b>	<b>-0.366</b>	<b>-0.916</b>	
	Maths	0.214 NS	-0.021 NS	<b>-0.827</b>	
	Science	0.321 NS	-0.133 NS	<b>-0.930</b>	
EAL	English	Not inc	Not inc	<b>0.582</b>	<b>-0.472</b>
	Maths	Not inc	Not inc	0.332NS	<b>-0.559</b>
	Science	Not inc	Not inc	<b>0.621</b>	<b>-1.244</b>

---

based on the whole cohort and do not take into account the specific attribute combinations of children.

We are interested in determining how other types of pupil fare compared to our reference pupil, hoping, thereby, to determine the probabilities for succeeding in the tests. The numbers (coefficients) in table 6 show how the reference category logits must be added to, in order to apply to other types of pupil, and to assess how their characteristics alter their chance of obtaining at least L4. The table has some cells containing 'Not Inc.'. This is because the model found was a sufficiently good representation of the data without the need for the additional complexity, which that term would introduce. Model diagnostics were used, in order to be satisfied that the model found was adequate.

The diagonal terms are called the 'main effects'. The signs of these are unsurprising. A positive sign means we add to the reference logit so that the probability of obtaining at least L4 is increased and vice versa for a negative sign. So for example being older increases the chances of obtaining L4 whilst being on FSM reduces it.

The off-diagonal terms are the 'interactions'. These are involved where the effect of two variables in combination is not the just the simple sum of the two separate effects. The off-diagonal term needs to be added to the two main effect terms in order to give the overall effect of those two variables in combination. As pointed out above the model was adequate and as such did not require any additional complexity beyond including some two-way interactions (i.e., there are no three-way interactions required to make additional adjustments when three variables are acting in combination.)

The numbers in bold type are statistically significant at  $p = 0.05$ , i.e. it is unlikely that chance would produce a number of this magnitude. The ones with NS beside them are not statistically significant and so there is not good evidence that they are other than zero. It was decided not to try to exclude these non-significant terms from the model on grounds of symmetry and also for comparison with the significant ones. The broad thrust of the conclusions are not too much affected by this choice.

We will now discuss the effects on achievement in each subject in turn, starting with the simplest situation to describe, that for Maths.

### **Accounting for Success in Maths**

Maths is the simplest situation. All the off diagonal terms are NS, which means there are no interaction effects. (Although if we needed to be more precise, we could include even non-significant effects when giving point estimates of the logits). The effect of being one year older (i.e., aged 12) is to raise the logit by 0.632.

There is no statistically significant female effect, i.e., boys and girls perform equally.

The FSM effect is significant and negative. It is of a large size, indicating that children on FSMs do poorly. At  $-0.827$  it counteracts the effects of being one year older. As we will see the same applies to the other two assessed subjects' main effects. However, we need to be careful in interpreting interactions, the terms for which are positive, suggesting that the effect of FSM is less severe for older children.

The effect for EAL is also significant and is similarly negative.

### **Accounting for Success in Science**

Science is the next simplest case in the sense that it only has one NS interaction.

The age effect is similar to maths. Again there is no gender effect. The FSM effect is large and negative, similar to that for Maths. The EAL effect is also negative and of greater magnitude than in the case of the other subjects.

Interestingly, the interaction term between FSM and EAL is positive. So whereas for FSM on its own the effect is  $-0.930$  and EAL on its own is  $-1.244$ , the effect in combination is  $-0.930 - 1.244 + 0.621 = -1.533$ . So the combined effect is worse than either alone but not as bad as just the sum of the main effects.

**Accounting for Success in English**

English is the most complicated as it has a significant gender main effect and three significant interactions. As before there is a significant age effect and an even bigger female effect—an 11-year-old girl is predicted to do better than a 12-year-old boy. The FSM effect is again large and negative. This has a positive interaction with age so that older children on FSM do considerably less badly than their younger counterparts. For example, whereas the effect of FSM on an 11-year-old boy without EAL is the basic  $-0.916$ , the effect for a 12-year-old boy without EAL is  $-0.916 + 0.433 + 0.608 = 0.125$ , i.e., only  $-0.916 + 0.608 = -0.308$  down on a non-FSM 12-year-old non-EAL boy.

For gender and FSM entitlement the interaction effect is negative so FSM takes away some of the advantage of being female. For a 12-year-old female on FSM the effect is the sum of the three main effects plus the two interactions.  $0.433 + 0.693 - 0.916 + 0.608 - 0.366 = 0.452$ , i.e., about the same as a 12-year-old boy without FSM, i.e., the last four terms of the sum roughly cancel.

The effect of EAL is not as negative with English as it is with the other two subjects. As in the case of Science, there is a positive interaction with FSM reducing the effect of both when they occur in combination. Indeed the interaction is large and positive which gives a combined effect of  $-0.916 - 0.472 + 0.582 = -0.806$  which is slightly less bad than the effect of FSM on its own. So the likelihood of achieving L4 in English is slightly greater if, given one is on FSM, one has English as an additional language. (However not too much should be read into small differences as the coefficients themselves have uncertainties associated with them.)

A particularly stark difference in predicted success is that of a 12-year-old girl without FSM or EAL compared to a 11-year-old boy on FSM but without EAL. The prediction of success for the girl is a logit of  $0.571 + 0.433 + 0.693 = 1.697$ , which gives a probability of success of 0.845. For the boy the logit is  $0.571 - 0.916 = -0.345$ , giving a probability of passing of only 0.415.

**Correlations Between Pupils' Subject Outcomes**

Correlations extracted from covariances and variances given from the multilevel model allow us to predict an individual pupil's chances of obtaining a L4 in other subjects, given success in one subject. The correlations in Table 7 are positive as expected. A pass in one subject is more likely to be associated with a pass in another. These are all quite precisely determined with a standard error of about 0.01.

Table 7 The correlations of individual success across subjects

	English	Maths	Science
English	1		
Maths	0.550	1	
Science	0.491	0.478	1

Table 8 School variability in subject success

	English	Maths	Science
Std deviation	0.594	0.491	0.869

Table 9 School correlation in subject success

	English	Maths	Science
English	1		
Maths	0.843	1	
Science	0.806	0.745	1

### Influence of Individual School on Subject Success

A school logit for passing each exam is essentially given from the model. In fact it is given as a school's departure above or below the overall logit for obtaining L4. This is known as the school level residual. The standard deviations of the variability between schools are displayed for each subject in table 8. The correlations are given in table 9.

Examining the standard deviations first, we see that science performance is more variable (S.D. = 0.869) between schools than either English or Maths (S.D. = 0.594 and 0.491, respectively).

The correlations for school success between the subjects are higher than that for individuals. School residuals for the exam subjects can be interesting in that they can show a few unusual cases, e.g., where a school is below average on one subject but is well above average on another.

### Discussion

Not surprisingly, we found that there is a positive effect for age in each subject. Previous research on SATs results has shown that older pupils scored better, on average, than younger children on all tested areas of the national curriculum (e.g., Schagen & Sainsbury, 1996).

Our findings on the impact gender are more noteworthy. Since 1998, the underachievement of boys has been represented in the media (Bright, 1998; Cassidy, 1999) and by the Office for Standards in Education (Dean, 1998) as hitting crisis proportions. As recently pointed out by Gorard *et al.* (2001) the pattern of boys' 'underachievement' is often seen as a uniform one with girls out-performing boys on all measures. Although based only on one LEA, albeit one of the largest outside London, our results show no significant gender effects in Science or Maths, but significant gender effects in English. This would suggest the need to temper the 'crisis' rhetoric and the need to explore achievement on a subject by subject basis. Findings by Gorard *et al.* (2001) and by Sammons *et al.* (1997) are in line with ours. Based on assessment results for Wales, including KS2 results, Gorard *et al.* found there were no large differences in attainment in Science and Maths between boys and girls at KS2 (or KS3). They concluded that this picture is remarkably constant for the two year period for which they had data. On the basis of our results, we can reasonably conclude that gender is not a problem in terms of SATs achievement in these two subjects in our LEA.



The different pattern we found prevailing in English is also mirrored in the Welsh study—they found a significant achievement gap between girls and boys for English at every Key Stage over a two year period and they found that this gap has widened in KS3. Future work with our cohorts will continue to explore the extent of the gap identified in English at KS2.

Being on FSMs negatively affects performance across all three assessed subjects. Being on FSMs is often taken as an indicator of social disadvantage or poverty. There is now extensive evidence on the way that poverty is linked with educational underachievement and failure, at both the pupil and the school level (e.g., Haveman & Wolfe, 1993; see Mortimore & Whitty, 1997, for a review) and more recent investigations support our findings here (Croll, 2002). But there is much less evidence of its impact over time—one recent study, using eligibility for FSM as an indicator, found that social disadvantage impacts negatively on the rate of improvement in examination results (Levacic & Woods, 2002). In other words those not on free school meals can expect to make more rapid progress than their counterparts who are on FSM. We propose to track the nature of this change for various cohorts of pupils within the LEA in future analyses. Of note is the point made by Mortimore and Whitty, that, while living standards overall increased over the period of their investigation (most of the nineties) the income gap between the rich and poor widened, resulting in greater numbers of children living in poverty. They concluded that even where schools can raise the achievement of their pupils, the relative achievements of children from poor and more affluent backgrounds are not likely to alter. These researchers note the silences surrounding issues of social disadvantage under the Conservative governments and expressed optimism that New Labour would ‘re-open this important public debate’. As Paul Croll observes in his recently published paper, such optimism was somewhat misplaced.

The previous point is also valid in the case of the effects of ethnicity on achievement. Our study did not have sufficiently accurate information about ethnic background to investigate the attainment of particular ethnic groups. However, given that the factor EAL impacts significantly and negatively on all three assessed areas, and that almost all those who have EAL belong to minority ethnic groups, it is likely that in our LEA the latter groups are disadvantaged accordingly. A review of the field commissioned by OFSTED in 1996 (Gillborn & Gipps, 1996) showed that there are considerable differences across ethnic groups and a more recent study by Strand (1999) based on KS1 data over the late nineties shows that the pattern hasn’t changed radically. The point made by Gillborn and Gipps (1996), that the question of race and equality of opportunity no longer has the the prominent position it once held, would seem to be still valid.

What is noteworthy in our findings is that, while not having English as a first language, impacts negatively on all three subjects, its effects differed across each subject. Its negative effect is largest in the case of Science and lowest in the case of English. One possible explanation for this effect difference in subjects, in our view, is the fact that the medium of the assessments is written English, which poses an additional challenge for those pupils whose fluency in that language may not be so good. They have to concentrate on the application of their conceptual knowledge or skills as well as how they are expressing themselves. They must demonstrate their scientific or mathematical understanding in what may be in reality a foreign language. The greater effect of EAL on Science than on Maths can be understood, in our view, on the grounds that the success in tests of the former requires more language competence. In contrast, the English tests invite and expect pupils to attend to matters of expression, to concentrate on language usage, and to prioritise language skill itself rather than conceptual knowledge. Unsurprisingly, other research, from inner London (e.g., Demie *et al.*, 1997) shows that children who are not fluent in English tend to do less well at all Key Stages than their counterparts who are fully fluent. While our research would

support the targeting of resources towards supporting those not fluent in English so they are able to fully access the curriculum, we recognise the need to incorporate more fine-grained information about levels of pupils' fluency into future analyses.

Our finding that Science achievement is much more variable across schools than either Mathematics or English merits some speculation. Although neither the national numeracy nor literacy strategies are statutory, it is highly likely in our view that their introduction into schools may be important in understanding the greater variation across schools in science. Both strategies are highly prescriptive in relation to (a) what is to be taught in both curricular areas, (b) how much time should be devoted to those areas, and (c) and perhaps more fundamentally, how these areas should be taught. Such circumscribed pedagogy is likely to narrow the variability in the learning experiences of pupils across schools which would, in turn, make for more similarity (than in the case of a subject not so circumscribed) in how pupils would perform in an examination.

### Conclusion

Previous research has tended to use less rigorous approaches to explore influences on achievement at the primary level. Although this research is based on a more rigorous methodological approach, we acknowledge its limitations. The absence of baseline prior attainment data precludes a longitudinal, 'value added' approach at this stage which would give some measures of school effectiveness. The nature of our sample—which is more ethnically diverse and which has a higher than average incidence of social disadvantage (as indicated by FSM eligibility) may mean these results may not be applicable to other geographical areas. For example, some 36% of the pupils in our study are on free school meals while the national average is half that at 18%. We are not in a position, nor was it our intention, to make claims about the national picture of SATs achievement at Key Stage 2.

We did not have information in the present study on pupil mobility—a factor found to impact negatively especially on maths (Strand, 2002). Analyses by Yang and Goldstein (1999) show that the number of previous schools and the length of time in the final school are associated with progress. Our ongoing work with the LEA, which is now logging this information, will take this into account in future analyses.

*Correspondence:* Dr Paul Marchant, Centre for Research and Graduate Studies, Leeds Metropolitan University, Leeds LS1 3HE, UK.

### Acknowledgements

Thanks are expressed for advice received from those at the Multilevel Models Project at the Institute of Education, London, particularly Min Yang.

### References

- AITKIN, M. & ZUZOVSKY, R. (1994) Multilevel interaction models and their use in the analysis of large-scale school effectiveness studies, *School Effectiveness and School Improvement* 5, 45–73.
- BRIGHT, M. (1998) The trouble with boys, *Observer*, 4 January.
- CASSIDY, S. (1999) Gender gap widens to a gulf, *Times Educational Supplement*, 29 January, 6.
- CROLL, P. (2002) Social deprivation, school-level achievement and special educational needs, *Educational Research*, 44, 1, 43–53.

- DEAN, C. (1998) Failing boys: public burden number one, *Times Educational Supplement*, 27 November, 1.
- DEMIE, F. (2001) Ethnic and gender differences in educational achievement and implications for school improvement strategies, *Educational Research*, 43, 1, 91–106.
- DEMIE, F., REID, A. & BUTLER, R. (1997) *Pupil Achievement and Ethnic Background: results of the analysis of the 1997 National Curriculum results*, Research and Statistics Unit, Southwark Education, London.
- GILLBORN, D. & GIPPS, C. (1996) *Recent Research on the Achievement of Ethnic Minority Pupils*, London, HMSO.
- GORARD, S., REES, G. & SALISBURY, J. (2001) Investigating the patterns of differential attainment of boys and girls at school, *British Educational Research Journal*, 27, 2, 125–139.
- HALL, K. & HARDING, A. (2002) Level descriptions and teacher assessment in England: towards a community of assessment practice, *Educational Research*, 44, 1, 1–16.
- HAVEMAN, R. & WOLFE, B. (1995) The determinants of children's attainments: a review of methods and findings, *Journal of Economic Literature*, 33, 1829–1878.
- JONES, K. & DUNCAN, C. (1998) Modelling context and heterogeneity: applying multilevel models. In: E. SCARBROUGH & E. TANENBAUM (eds) *Research Strategies for Social Scientists*, Oxford, Oxford University Press.
- LEVACIC, R. & WOODS, P. (2002) Raising school performance in the league tables (part 1): disentangling the effects of social disadvantage, *British Educational Research Journal*, 28, 2, 207–226.
- MORTIMORE, P. & WHITTY, G. (1997) *Can School Improvement Overcome the Effects of Disadvantage?* London, University of London Institute of Education.
- PATERSON, L. & GOLDSTEIN, H. (1991) New statistical methods for analysing social structures: an introduction to multilevel models, *British Educational Research Journal*, 17, 4, 387–393.
- QUALIFICATIONS AND CURRICULUM AUTHORITY (QCA) & DEPARTMENT OF EDUCATION AND EMPLOYMENT (DfEE) (1999) *The National Curriculum: handbook for primary teachers in England*, London, HMSO.
- SAMMONS, P., WEST, A. & HIND, A. (1997) Accounting for variations in pupil attainment at the end of key stage 1, *British Educational Research Journal*, 23, 4, 489–512.
- SCHAGEN, I. & SAINSBURY, M. (1996) Multilevel analysis of the key stage 1 National Curriculum Assessment data in 1995, *Oxford Review of Education*, 22, 3, 265–272.
- 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.
- YANG, M. & GOLDSTEIN, H. (1999) The use of assessment data for school improvement purposes, *Oxford Review of Education*, 25, 4, 469–484.