# Changes and improvements in schools' effectiveness: trends over five years

John Gray, Harvey Goldstein and David Jesson

**ABSTRACT**

It is widely assumed that some schools improve more rapidly than others. However, unlike the well-established finding that schools differ in their effectiveness, evidence that schools improve at different rates is sparse.

Using data on pupils' examination results and prior achievements from five cohorts of pupils passing through over 30 schools in one LEA, the study considers three questions. First, the extent to which some schools improve more rapidly than others. Second, whether certain 'types' of school are more likely to improve. And third, whether there were any 'strategies' which substantial numbers of schools employed to bring about improvement.

A multilevel model was employed to generate estimates of schools' 'effectiveness' and 'improvement over time'. The analysis showed that around 1 in 7 schools 'improved rapidly' over the 5-year period. Both 'less effective' and 'more effective' schools' improved. However, differences in schools' effectiveness remained substantial when compared with the extent of improvement; it would take several years, on this evidence, for a 'less effective' school to move into the pack and as long again for it to pull ahead. One reason why some schools were improving more rapidly than others could be attributed to the fact that they had increased the average number of exam subjects their pupils were entered for above the rate of increase across the whole sample. Other factors not explored in this study are also likely to have been influential.

**Keywords:** school effectiveness; school improvement; value-added; examination results; multilevel modelling; trends in performance

Harvey Goldstein is Professor of Statistics at the Institute of Education, University of London; John Gray is Professor and Director of Research at Homerton College, Cambridge; David Jesson is a Senior Lecturer at the University of Sheffield and Visiting Professor in the Centre for Performance Evaluation and Resource Management, Department of Economics, University of York.

*Research Papers in Education* 11(1) 1996, pp. 35–51 

## INTRODUCTION

It is widely recognized that school improvement takes place over long periods of time, but the implications of these time periods for research on school effectiveness and school improvement are still in the process of being understood. Crucially what is required are studies of how schools change over fairly extended periods of time. There are practical reasons why the time-scales of most research on school effectiveness have been all too brief. Considerable resources are required to study even a single cohort of pupils as they pass through a stage of schooling. To study how schools change, however, requires several cohorts of pupils.

Few of the major studies of school effectiveness to date, whatever their aspirations, have been in a position to say much about the longer-term processes of initiating school improvement and building greater effectiveness. Teddlie and Stringfield's (1993) 10-year study of effective and ineffective elementary schools in Louisiana is the major exception to this general picture. We know quite a lot about what distinguishes a more effective school from a less effective one but rather less about how such schools become more effective or, for that matter, less effective — unless, that is, we assume that the processes are much the same.

Most studies of school improvement have suffered from similar limitations. In his authoritative review of the research literature on change, Fullan (1991) concludes that: 'significant change in the form of implementing specific innovations can be expected to take a *minimum of two or three years*; bringing about institutional reforms can take *five or more years*' (p.106, our emphases). Changes in schools' effectiveness would almost certainly fall into this latter category.

If one takes Fullan's assessment of the time dimension seriously then it rapidly becomes clear that the typical improvement study is likely to be about schools' early experiences. Most studies simply come to a halt before the time when the outcomes of the changes could realistically be expected to have occurred. They can sound optimistic about the prospects and they can point to some of the likely outcomes but the reader has to take much on trust and assume that the best-laid plans will be carried through.

## KEY QUESTIONS

The many implications of bringing together the insights from the two research traditions of school effectiveness and school improvement are discussed more fully elsewhere (see Gray *et al.*, 1996). Of particular interest in the present study are some key questions about the nature and extent of school improvement.

1. Do some schools improve more rapidly than others and, if they do, by how much? Related to this is the question of how stable schools' performances are over time? For there to be 'improvement', of course, there must also be some 'instability' but what is the relative importance and contribution of each?
2. Is there any tendency for certain types of school to improve more rapidly than others? Do those which are already 'more effective', for example, stretch their lead still further? And what of the fate of those which are markedly 'less effective'? Do they fall still further behind?

3. Are there any 'strategies' which substantial numbers of schools have been using to bring about improvement? Each school is likely to have its own 'natural history' of change and previous research might lead one to expect a wide range of approaches. But are there any that seem to hold across institutions?

Similar questions, of course, helped to structure pioneering research on school effectiveness. Expressed in these terms such issues may seem rather obvious, but unfortunately, to date, there is no comparable body of school improvement research to which to turn. The task of carefully linking processes to outcomes has only occasionally been undertaken. This is not to devalue the important findings that have emerged from research on school improvement but merely to observe that the largely qualitative approaches which have dominated this paradigm have not, until recently, accorded evidence about outcomes as much salience as that relating to processes. In our view to be able to assert not only that schools do change and improve but by how much and over what periods of time is an important part of the research agenda on which subsequent research (and reassessments of the available evidence) can then be built. We note also that others are beginning to agree (see, for example, Fullan and Hargreaves, 1992; Hopkins *et al.*, 1994).

In comparing the extent to which schools have improved over time there is, of course, a pitfall of which one needs to be aware. The most obvious reason why one school might have improved in some particular area when compared with another which has made little or no progress is that it strove to do so. As Barber (1994) has reported, in a national survey of some 60 English improvement projects only one area was 'intended for improvement' by more than three-quarters of the projects participating. This was, not surprisingly, the area of 'pupil attainment' (85 per cent of projects reported this as a target). The next most frequently mentioned areas were 'pupil/teacher expectations' (mentioned by 72 per cent), 'pupil self-esteem' (67 per cent) and 'parental involvement' (60 per cent). In the present case the outcome measure of interest is the schools' examination results. Whilst it would be hard, on the above evidence, to argue that a school would not be interested in further improving the exam performances of its pupils, clearly different schools might make different decisions about the relative importance of competing objectives for improvement.

## THE POLICY CONTEXT

There is currently considerable interest in policy circles in ways of fostering school improvement. In the summer of 1994, for example, a Working Party of the School Curriculum and Assessment Authority on so-called value-added performance indicators suggested that a 'school improvement index' might be published alongside schools' 'raw' examination results (SCAA, 1994). This would, it was hoped, provide a way of calculating the extent of progress schools had made from their original starting points (say, 3 years earlier). A short while after schools' 1994 examination results had been published the Office for Standards in Education (OFSTED) singled out some 50 'improving' schools for praise. These schools had all apparently made substantial progress in terms of getting increased percentages of their pupils over certain exam hurdles. OFSTED has also published three case studies of 'improving schools' (OFSTED, 1994) and has further work in hand. Finally, in the middle of 1995, the Department for Education announced funding for a

major programme of school improvement initiatives. The concern to understand more about the nature and extent of school improvement can rarely have been higher. As Brown (1995) has argued, however, moves to turn research on school effectiveness and improvement into policy 'recipes' need to be resisted. There are likely to be several routes.

Such developments take place against a background of rising expectations about the performance of the educational system and individual schools within it. The introduction of so-called league tables may also have encouraged schools to 'improve their performance' by one means or another. One contributory factor (and an important one at that) is likely to be the extent to which they have managed to attract more able pupils. None of the public attempts to date to measure or assess the extent of improvement has tried to take such considerations systematically into account. Nor, for that matter, have most researchers of school improvement been in a position to do so. The assumption has tended to be that one year's intake is pretty much like the next year's in terms of prior achievements. The evidence we have available suggests that this is not always the case — intakes may fluctuate and the opportunity for parents to choose schools may lead to greater changes in their intakes' achievements than occurred in the past. The crucial long-term question, of course, is whether schools have 'improved in their effectiveness'. In other words, given the kinds of pupils they have recruited, are they succeeding in helping them to achieve more? And are pupils in this year's cohort achieving more than 'similar' pupils achieved last year?

A note of caution needs to be sounded at this stage about the language of school improvement. Effectiveness and improvement are relative terms and may be brought about in different ways. We shall follow the convention of referring to schools as being 'more' or 'less' effective and improving 'more rapidly' or 'more slowly'. As Reynolds (1996) has observed, however, there is a 'dark side' to the equation which has been much less explored. Researchers know little enough about how 'ineffective' schools function and even less about how they begin to deteriorate. In some respects it may be a zero-sum game. To couch explanations largely or exclusively in positive, action-oriented terms therefore, as much of the school effectiveness and improvement literature tends to do, may be to blind ourselves to the realities underlying change processes. Avoiding the negative may be as important as accentuating the positive.

## THE DATA

The data for this study were drawn from 34 secondary schools in one LEA (data on a further school were incomplete for one of the 5 years). They covered five successive cohorts of young people as they passed through the schools. The first cohort entered secondary school (year 7) in the autumn of 1985; this group took their GCSE examinations (in year 11 aged 16+) in the summer of 1990. The fifth cohort entered in the autumn of 1989 and took their exams in 1994 (see Table 1). Together they cover pupils' experiences of schooling for the better part of a decade.

The data available on each pupil concerned a number of factors and included a common measure of prior attainment at the point of entry to secondary school (based on the Reading Test DE produced by the National Foundation for Educational Research), their subsequent performance in GCSE exams (which was used, in line with recent practice in research on school effectiveness in Britain, to calculate an exam points score with 7

**Table 1: Contextual characteristics of the pupil sample**

| | 1990 | 1991 | 1992 | 1993 | 1994 |
|---|---|---|---|---|---|
| Prior achievement | 102.1 | 101.5 | 102.2 | 101.9 | 101.9 |
| SD | 12.4 | 12.7 | 12.9 | 12.4 | 12.4 |
| Gender (% female) (male = 0; female = 1) | 48.4 | 50.7 | 51.8 | 49.1 | 49.1 |
| Percentage on free school meals* | n/a | 7.0 | 6.7 | 9.1 | 9.1 |
| Percentage of pupils statemented* | n/a | 1.7 | 2.0 | 2.4 | 2.4 |

* These two variables were not employed in the analysis reported here as data were only available for four of the five years.

points being awarded for an A grade, 6 for a B down to 1 for a G). Their gender was recorded. Measures of poverty (entitlement to free school meals) and whether the pupil was statemented for special needs were also available for the second to fifth cohorts but were not used in the analyses covering all five cohorts. Some analyses were also undertaken on pupils' performances in separate subjects such as English and mathematics but are not reported here.

For the first cohort, information was based on a one-in-five random sample of pupils. For the four subsequent cohorts the full populations were used. A retrospective matching procedure was employed; for a variety of reasons, however, it was not possible to come up with prior attainment measures for all those pupils for whom GCSE exam results were available. The matching process resulted in a total sample with complete data on 14,965 pupils across the five cohorts. The contextual characteristics reported in Table 1 suggest that the overall population of pupils entering the schools over the 5 years did not change markedly from year to year.

## PATTERNS OF IMPROVEMENT

Figure 1 presents the aggregate exam results of the 34 schools across the five cohorts. It will be immediately obvious that the average exam points score per pupil improved over this 5-year period, as indeed it did in many other schools and areas of the country.

For the purposes of presenting the lines in the figure the results of a selective school (whose pupils, not surprisingly, scored considerably above the other schools in terms of exam results) have been omitted from this and the subsequent graph — but not from the analyses. The first cohort averaged 28.5 points per pupil across the sample as a whole (with a standard deviation of 15.4 points). This rose a little amongst the second cohort

**Table 2: Changes in pupils' average exam point scores 1990–1994 (whole cohort)**

| | 1990 | 1991 | 1992 | 1993 | 1994 |
|---|---|---|---|---|---|
| Average score per pupil | 28.5 | 29.1 | 34.2 | 35.9 | 37.6 |
| SD | 15.4 | 15.9 | 16.9 | 16.5 | 17.2 |

to 29.1 points. The third cohort was the first to have its results reported as part of the Government's decision to publish all schools' exam performances; the average at this point rose to 34.2 points. It continued to rise further with the fourth cohort (to 35.9 points) and by the end of the study had reached 37.6 points amongst the fifth cohort (with a standard deviation of 17.2 points) (see Table 2). There are several ways in which this overall improvement of some 9 points might be expressed in terms of exam points scores. It is ,equivalent, for example, to the average pupil securing a C grade award in one new subject and raising his or her performance in four others by a grade (say from a D to C or an E to a D in each).

Looking at the results from individual schools it is clear that virtually all improved to some extent. The general impression is of an 'upward march' over the whole period. In every case the average exam points score per pupil of the fifth cohort was above that of

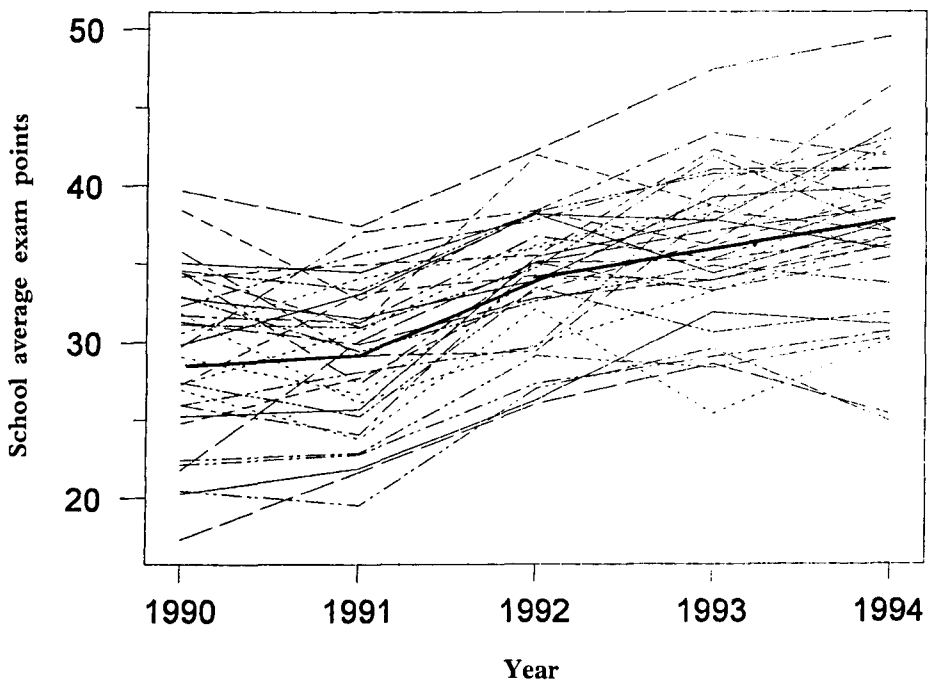


**Fig. 1: Exam points per pupil over 5 years — changes in schools' performances over time.**

the first cohort. This pattern was replicated nationally. In a number of cases, however, the extent of improvement was rather modest and might prompt questions about whether there had really been any improvement; in three schools, for example, the upward shift was less than 3 points per pupil against an average for the whole sample of 9. Whilst these schools might be said to have 'improved' their performance over that in previous years it is evident that the extent of 'improvement' can also be judged in relative terms; others, on this basis, appear to have achieved rather more. In other cases the year-on-year pattern was rather variable; the 'march' was not always steady. We are also aware, of course, that there is some debate about whether changes in pupils' exam performances indicate that underlying educational standards are rising.

There are several substantive reasons why a school's performance might change from year to year. Two are of particular interest in the present context. The most obvious one is that the prior attainments of the schools' intakes changed substantially from year to year. On this argument the schools which had 'improved' would only have appeared to have 'got better'; the underlying factor would have been a 'better' intake. Inspection of the correlations between schools' intakes across any 2 years indicates very high levels of correlation (well above 0.9 in all but one case; data not shown). In other words most schools had similarly attaining intakes from one year to the next (see Fig. 2). Inspection of the intakes of individual schools, however, reveals some changes within this stable pattern. One of the schools, for example, whose results improved least had a deteriorating intake in terms of performance over the five cohorts; the year-on-year correlations between intakes were high but the overall trend was nonetheless downwards. On the other hand,

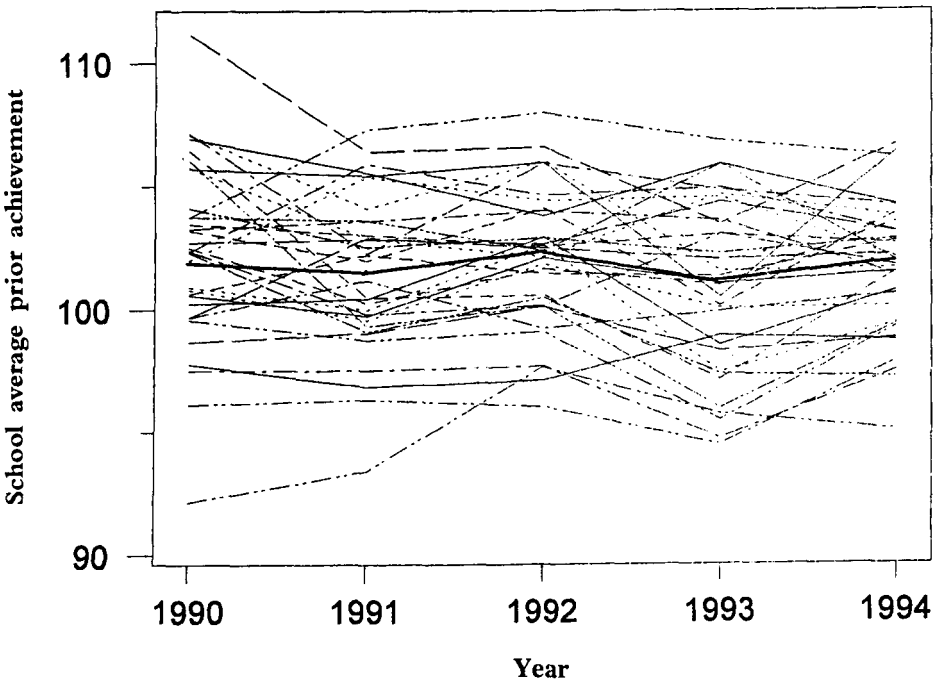


**Fig. 2: Prior achievement score per pupil over 5 years — changes in schools' intakes over time.**

the school whose results improved most over the same period had an essentially stable intake.

As a general rule, in the vast majority of schools, the performance of the previous year's intake was a good guide to that of the current one. There were, however, a number of exceptions to this general rule, sufficient to suggest that a more sensitive analysis, which deliberately took school-by-school differences in intakes over time into account, would be required.

## MODELLING IMPROVEMENTS IN SCHOOLS' EFFECTIVENESS

The main focus in the analyses which follow is on the extent to which schools 'improved in their effectiveness'; 'improvement' is, of course, a relative concept and may, in certain instances, represent a deterioration. Two stages are involved in estimating such effects. The first is to establish each institution's 'effectiveness' with the first cohort of its pupils; some schools will doubtless turn out to be more effective than others. The second step is to repeat the exercise for each of the schools for each of the subsequent cohorts. Two features of the resultant estimates are of interest. First, the extent to which they provide evidence that a school was more or less 'effective' at the start of the period under review; and second the extent to which such estimates of effectiveness were changing over time.

Multilevel modelling has become an increasingly widely employed technique for studying data of the hierarchically structured kind available here and was consequently employed in all the analyses which follow (see Goldstein, 1995, for an extended discussion). Following the usual procedures in this kind of work, the outcome measure for each pupil (their exam points score) was transformed to a normal score scale.

### Two-level model

The modelling process was initially conceived as a two-level model in which pupils (level 1) were nested in schools (level 2). Several models were tried out and in the fixed part of the model a mean was fitted for each year. Gender was also fitted and at level 1 (between pupils) different variances for males and females were allowed. Table 3 shows a model where the effect for each school for each year was allowed to vary.

The effects of the basic explanatory factors (prior achievement, gender and year) which were entered into the fixed part of the model are shown in the lower half of Table 3. All three made statistically significant contributions (twice their standard errors); several interaction terms also proved significant.

Not surprisingly, pupils' prior achievements made the strongest contribution to the prediction of pupils' subsequent exam performances. Squared and cubic terms relating to this factor also made modest additional contributions to the prediction. Table 3 shows that gender and year had effects as well. Females achieved at higher levels than males and general levels of performance improved over time. The consequences of these two effects are presented in Table 4, which shows, by way of example, that the estimated score for males and females of average prior attainment rose in each successive year.

Several interaction terms made modest additional contributions. Females with higher prior achievements, for example, attained outcomes additional to those due to these factors

**Table 3: Between-school variances (diagonal) and correlations across years**

| | 1990 | 1991 | 1992 | 1993 | 1994 |
|---|---|---|---|---|---|
| *Year* | | | | | |
| 1990 | 0.023 | | | | |
| 1991 | 0.88 | 0.027 | | | |
| 1992 | 0.75 | 0.84 | 0.026 | | |
| 1993 | 0.52 | 0.77 | 0.86 | 0.042 | |
| 1994 | 0.56 | 0.71 | 0.86 | 0.81 | 0.051 |
| *Between-student variances* | | | | | |
| Intercept | 0.384 | (0.034) | | | |
| Int/female | –0.0087 | (0.005) | | | |
| Int/year | 0.015 | (0.002) | | | |

| Fixed coefficients | Estimate | Standard error |
|---|---|---|
| Intercept | –0.465 | |
| Female | 0.23 | 0.01 |
| Year 1991 | 0.07 | 0.03 |
| Year 1992 | 0.34 | 0.03 |
| Year 1993 | 0.51 | 0.04 |
| Year 1994 | 0.56 | 0.04 |
| Prior achievement | 0.049 | 0.001 |
| Prior achievement$^{(2)}$ | 0.00012 | 0.000054 |
| Prior achievement$^{(3)}$ | –0.0000101 | 0.0000013 |
| Female* year | –0.008 | 0.004 |
| Female* prior achievement | 0.0024 | 0.0009 |
| Year* prior achievement | 0.0023 | 0.0004 |
| Year* (Prior achievement)$^{(2)}$ | –0.000048 | 0.00002 |
| Likelihood ratio $\chi^2$ [–2 ln (likelihood) = 29974.0] | | |

Level 1 standard errors in parentheses. Interaction terms in fixed part involving 'year' use this as a continuous variable (0–4).
*: multipled by

acting independently. There was also some very modest evidence that the influence of prior achievement itself increased with year

Two features of the between-school variance are of particular note. First, there is some limited evidence that this increased over the course of the study. The top half of Table 3 shows the 'between-school' residual variance for each of the 5 years, after the factors identified in the model had been taken into account. Whilst relatively small in size, this

**Table 4: Estimated (N) scores from the multilevel model for male and female pupils of average prior achievement**

| | 1990 | 1991 | 1992 | 1993 | 1994 |
|---|---|---|---|---|---|
| Male | −0.465 | −0.395 | −0.125 | +0.045 | +0.095 |
| Female | −0.235 | −0.165 | +0.105 | +0.275 | +0.325 |

measure increased over the 5 years. In the first year it was around 5 per cent (derived from the evidence in Table 3); by the fifth year it had risen to around 8 per cent.

Second, there was strong evidence of changes in schools' 'effectiveness' over time (see Table 3). The correlation between the residuals in year 1 and year 2 was 0.88. As the time period between comparison points lengthened, however, the correlations reduced. Between year 1 and year 5, for example, it had fallen to 0.56 (see Table 3). Knowing how 'effective' a school was in year 1, therefore, was only a partial guide to how 'effective' it would appear to be several years later. We note that a similar figure over a 5-year period was reported in a study of the stability of performance amongst subject departments by Fitz-Gibbon *et al.* (1989).

Table 3 also shows evidence of greater variance between pupils (level 1) over time, after taking account of differences due to factors in the fixed part of the model (see the int/year coefficient for the between-student variance in the table). This is in line with the overall increase in the variability of pupils' exam points scores signalled earlier in Table 2.

## Three-level model

The second stage of the modelling process was to fit a three-level model. In this analysis the pupils represent the level 1 units, grouped by year within school (at level 2) with the schools themselves being the level 3 units. At level 3 both year and prior achievement effects were allowed to vary randomly across schools, as well as the intercept. The results of the first three-level analysis are shown in Table 5.

The fixed effects in Table 5 are very similar indeed to those presented earlier in Table 3. The main interest lies in the overall structure of the variance. The table shows how the variation was allocated between the various levels. The between-schools component was 0.022 whilst the within-schools–between-years component was 0.0044. In the context of the present paper the most interesting feature of these estimates is the size of the between-year variation as a component of the between-schools variation. The evidence indicates that schools were fairly stable in their 'effectiveness' from year to year. Only about one-sixth (17 per cent) of the between-schools variation could be attributed to the between-years component. In other words there was a fair degree of stability in the year-on-year estimates of schools' effectiveness; on the other hand there was some modest evidence of change. In combination these findings suggest that schools which markedly

**Table 5: Between-school variances for intercept and coefficients of year and prior achievement with between-year variance for intercept (correlation off-diagonal)**

| | Intercept | Year | Prior achievement |
|---|---|---|---|
| *Between school* | | | |
| Intercept | 0.022 | | |
| Year | -0.07 | 0.0018 | |
| Prior achievement | 0.27 | 0.44 | 0.00005 |
| *Between year* | | | |
| Intercept | 0.0044 | | |
| *Between student* | | | |
| Intercept | 0.37 (0.01) | | |
| Int/female | -0.0065 (0.005) | | |
| Int/year | 0.016 (0.002) | | |

| Fixed coefficients | Estimate | Standard Error |
|---|---|---|
| Intercept | -0.46 | |
| Female | 0.23 | 0.01 |
| Year 1991 | 0.06 | 0.03 |
| Year 1992 | 0.34 | 0.03 |
| Year 1993 | 0.50 | 0.04 |
| Year 1994 | 0.55 | 0.04 |
| Prior achievement | 0.050 | 0.002 |
| Prior achievement$^{(2)}$ | 0.000052 | 0.000054 |
| Prior achievement$^{(3)}$ | -0.0000101 | 0.0000013 |
| Female* year | -0.008 | 0.004 |
| Female* prior achievement | 0.0025 | 0.0009 |
| Year* prior achievement | 0.0020 | 0.0004 |
| Year* (prior achievement)$^{(2)}$ | -0.000041 | 0.00002 |
| Likelihood ratio $\chi^2$ [-2 ln (likelihood) = 29797.2] | | |

*: multiplied by

changed in their effectiveness over the 5-year period of the study are likely to be in fairly short supply.

Table 6 presents a final refinement of the model in Table 5. Again the fixed effects are very similar to those presented earlier, as are the level 3 factors (year and prior achievement). In addition, however, gender and prior achievement are treated as random at level 2. This model yields rather similar estimates of the variance structure (between-schools 0.019 and

**Table 6: Between-school variances for intercept and coefficients of year and prior achievement with between-year variance for intercept (correlation off-diagonal) and further random coefficients between years**

| | Intercept | Year | Prior achievement |
|---|---|---|---|
| *Between school* | | | |
| Intercept | 0.019 | | |
| Year | 0.029 | 0.0016 | |
| Prior achievement | 0.26 | 0.49 | 0.000046 |
| *Between year* | | | |
| Intercept | 0.0034 | | |
| Prior achievement | 0.44 | 0.000013 | |
| Female | −0.13 | 0.05 | 0.0079 |
| *Between student* | | | |
| Intercept | 0.37 (0.01) | | |
| Int/female | −0.0074 (0.005) | | |
| Int/year | 0.016 (0.002) | | |

| Fixed coefficients | Estimate | Standard Error |
|---|---|---|
| Intercept | −0.45 | |
| Female | 0.23 | 0.01 |
| Year 1991 | 0.06 | 0.03 |
| Year 1992 | 0.33 | 0.03 |
| Year 1993 | 0.50 | 0.04 |
| Year 1994 | 0.55 | 0.04 |
| Prior achievement | 0.050 | 0.002 |
| Prior achievement$^{(2)}$ | 0.000035 | 0.000054 |
| Prior achievement$^{(3)}$ | −0.0000101 | 0.0000013 |
| Female* year | −0.009 | 0.004 |
| Female* prior achievement | 0.0026 | 0.0009 |
| Year* prior achievement | 0.0020 | 0.0004 |
| Year* (prior achievement)$^{(2)}$ | −0.000032 | 0.00002 |
| Likelihood ratio $\chi^2$ [−2 ln (likelihood) = 29769.8] | | |

*: multiplied by

within-schools–between-years 0.0034) indicating that the within-schools–between-years component accounted for around one-seventh (15 per cent) of the total.

## INTERPRETING THE RESULTS OF INDIVIDUAL SCHOOLS

What did the picture look like in relation to individual schools? Figure 3 shows the modelled estimates of change plotted in relation to effectiveness. The vertical axis describes each school's effectiveness (the 'intercept') in the initial year of the study (1990); the higher up the scale a school is, the more effective it was. The horizontal axis provides the additional estimate of each school's 'improvement' (the year 'slope') over the 5-year period; the further it is to the right of the scale the more it was improving its effectiveness over time. These estimates, it should perhaps be stressed, are for the average pupil in each school; the picture would be slightly more complex if the performances of pupils who were at either the upper or lower ends of the achievement scale in each school were to be presented.

Both the 'effectiveness' and 'improvement' estimates for each school need to be interpreted in light of their inherent uncertainty. Each estimate has an associated standard error (data not shown). Only when such estimates are significantly different from zero can one be confident that the school is really different from the population as a whole. At the same time 'bands of uncertainty' (as they are called) can be constructed around the estimates for individual schools. Only when these do not overlap can one be confident that two individual schools are really different from each other (see Goldstein and Healy, 1995, for a fuller discussion).

Inspection of the detailed school-level residuals and standard errors for each school in the study revealed some notable findings (data not shown). Overall, of the 34 schools in the study, six had positive 'intercept' residuals; they were, in other words, significantly more effective than other schools. In similar vein, five schools had statistically significant negative 'intercept' residuals; they were significantly less effective than other schools. These estimates of the effectiveness of individual institutions are comparable to those generated by earlier research on school effectiveness which has suggested that around two-thirds to three-quarters of institutions perform around the levels to be expected from information about their intakes with a one-quarter to a one-third falling outside these limits.

At the same time there were ten schools in the study whose 'year' slope estimates were of statistical significance (data not shown); they were divided equally between those which were improving fairly rapidly and those which were improving more slowly.

### Effectiveness and improvement

It will be clear from this discussion that schools' 'performances' can be judged in terms of both of these dimensions — effectiveness (the 'intercept' axis) and improvement (the 'year' axis). We have therefore divided Fig. 3 into four quadrants and identified ten schools for particular comment.

The upper right quadrant of Fig. 3, for example, contains schools which were initially more effective and which subsequently improved more rapidly. Three of the schools fell into this quadrant. Two of them were of average effectiveness to begin with but improved

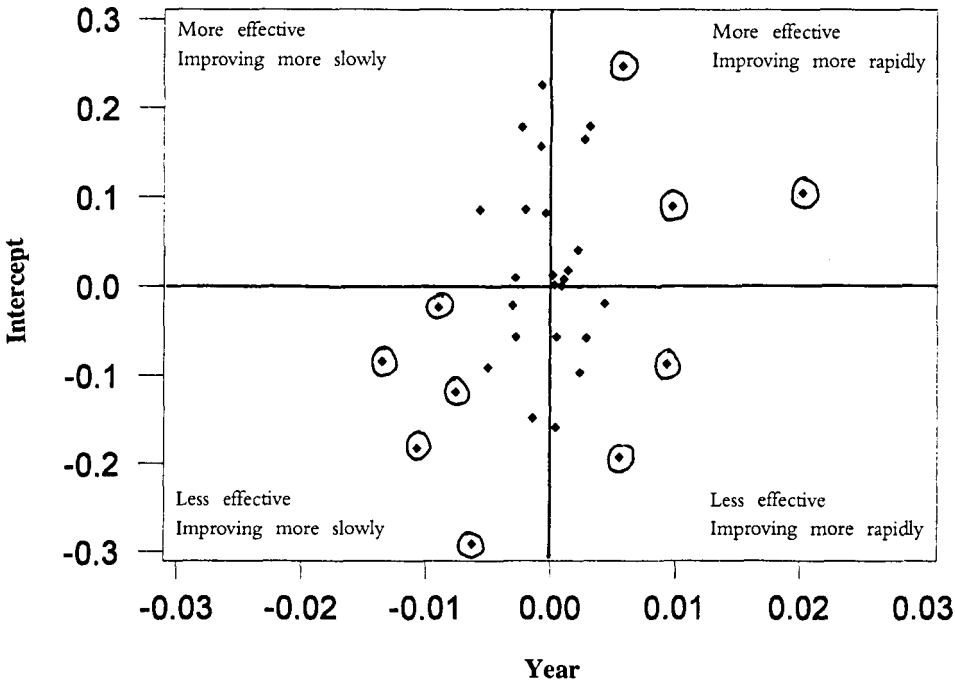


**Fig. 3: Intercept by year school level residuals.**

fairly rapidly from there; one was initially more effective and also achieved some further improvement. The lower left quadrant of Figure 3 comprises schools which were initially less effective and which were improving more slowly in comparison with the others. Of the five schools we have identified in this quadrant, one was of average effectiveness at the start of the study and one was less effective whilst three fell somewhere in between these two in terms of their effectiveness. In the bottom right-hand quadrant are two schools which improved more rapidly over the 5 years of the study. One of these was of around average effectiveness at the beginning; the other was a less effective school which had begun to improve. For completeness one should probably look at the upper left-hand quadrant as well which contains schools that were more effective but improving more slowly; interestingly, however, none of the schools in this quadrant could be singled out for particular comment in terms of the criteria we have adopted.

Figure 3 suggests that improvement can come from a number of starting points. Both more effective and less effective schools improved fairly rapidly over time (see the upper and lower right-hand quadrants). Conversely, more and less effective schools also improved more slowly. There would appear, however, to be an imbalance here between the upper and lower left-hand quadrants. Schools which were already less effective seem to have been more likely to have been found in the lower quadrant rather than the upper one. The least effective school in the sample, for example, was one of those which improved more slowly. On the other hand, amongst the other less effective schools it will be noted that we have singled out two which diverged — one improved more rapidly, the other less so.

Overall the relationship between effectiveness and improvement was neither strong nor very clear-cut (the correlation was a rather modest 0.26). Questions about whether 'improvement' is a steady year-by-year process or one which goes in fits and starts also remain open. Data on additional cohorts would be necessary for clearer patterns to emerge. Further replications of these findings in other contexts would also be desirable before firmer conclusions could be reached.

## THE EFFECTS OF CHANGES IN ENTRY POLICY

There are many reasons why schools' performances may have changed over time. Most of these would require extensive exploration in school settings to place on a firmer footing and it is our intention in the near future to undertake such investigations. One factor, however, can be partly considered from the data presently available.

It is a commonplace observation that schools' performances are affected by their entry policies. National trends were reflected in our local samples. Over the period covered by this study the average number of GCSE examinations entered by the average pupil in the sample rose from 7.4 in 1990 to 8.6 in 1994. Our evidence indicates, however, that the rise was essentially taken in a single step. In 1991 the average number of entries per pupil was 7.5; in 1992 it jumped by an average of just under one exam entry per pupil to 8.4. To what extent was this factor contributing to the changes?

Despite the frequency with which the entry-policy factor is mentioned in discussions with practitioners it has proved rather difficult to demonstrate the effects. Some of the problems are discussed in Blakey and Heath (1992). What most researchers have had available is a measure of the number of examinations each member of their sample has been entered for. Unfortunately, the number of entries per pupil is not the same as the schools' entry policies because all schools, whatever their distinctive approaches, tend to enter more able pupils for larger numbers of examinations. The total number of entries becomes, in practice, the school's crude estimate of the pupil's probable success in the forthcoming examinations. Since schools usually base entry decisions to some extent on pupils' prior performances in 'mock' examinations, the 'total number of entries measure' serves in a similar way to any prior achievement controls which might be available, although being based on more recent information it may appear to be a better predictor.

In brief, getting at the 'entry policy' factor is not as straightforward as is sometimes believed. Schools' overall practices (their policies) are built up from numerous decisions about individual pupils which it would be hard to recover without detailed discussions about each in turn. What the present study can model is the changes in those policies across whole schools which came about, for whatever reasons, during the 5-year period under review. The variable which was constructed related to these changes. For each school in the sample the average number of exams entered was calculated. The value for the first year was subtracted from each average for each school. The typical school in the study, therefore, would have had a value of just over one on this measure. Four schools in the sample, however, had upped their average entries per pupil by over two.

The model reported in Table 6 was rerun with the entry change variable as well. This variable turned out to make a statistically significant contribution to the equation taking account of all the other factors (data not shown). In other words there was a tendency for schools which changed (or were in a position to change) their entry policies, by

putting their pupils in for more exams, to improve their overall performances over time more than those which changed them (for whatever reasons) more modestly or more slowly.

## CONCLUSIONS

The availability of data on five cohorts of pupils passing through the secondary stages of schooling offers one of the first opportunities to explore the extent to which schools change in their effectiveness over time. There is clear evidence from the study that schools differed. Some improved rapidly, others more slowly. The extent to which schools appeared to improve, however, was fairly modest compared with the extent to which they initially differed in their effectiveness. Our evidence suggests that it would take several years for a relatively ineffective school to get into the pack of schools deemed to be initially of average effectiveness and as long again for them to pull ahead. For them to do so would also imply that they had somehow acquired the ability to improve consistently from year to year. Very few of the schools in the study appeared as yet to be in this position.

A second concern of the present study was to see whether there were any clear patterns regarding the schools which improved. Were those which were already more effective likely to pull still further ahead? Or were there also signs of comparable changes amongst less effective schools? The evidence on this point is more difficult to interpret and one needs to be aware of the small numbers upon which any conclusions about schools in this study can be based. Both more and less effective schools were present amongst those which improved more rapidly; amongst the group which improved more slowly there were more schools that were initially less effective. However, the overall correlation between the 'effectiveness' and 'improvement' dimensions was sufficiently low for the question to remain a moot point.

There was one particular 'strategy' schools in the study which were improving 'more rapidly' had adopted at some (usually earlier) point over the 5 years. This was the decision to increase the average number of examination subjects for which their pupils were entered. The typical school in the study upped the average number of entries per pupil by one exam subject; there appeared to be some pay-off in terms of 'improvement' to those schools which went beyond this. Whether this was essentially a 'one-off' strategy (concentrated perhaps on the higher-achieving pupils) or one which could be further developed remains to be seen. Whilst it was clearly an important contributor to the changes which were observed, it also fell a good way short of explaining all or even most of them. Whether many other common factors will be found remains to be seen. The evidence from the available literature suggests that the roads to improvement may be more varied (see Gray and Wilcox, 1995, for fuller reviews).

The fields of research on school effectiveness and school improvement have, until recently, been pursued by researchers with different methodological and substantive orientations. There have been several promising signs in the last 2 or 3 years of a greater concern to understand the problems to be faced in integrating the two traditions. The evidence in this paper suggests that this could be well worthwhile.

## ACKNOWLEDGEMENTS

The analyses reported here were supported by a grant to John Gray, David Hopkins and David Reynolds from the Economic and Social Research Council for 'A longitudinal study of school change and improvement' (the Improving Schools Project)(R000235864).

## REFERENCES

BARBER, M. (1994). Urban Education Initiatives: The National Pattern. A report for the Office for Standards in Education, Keele University.

BLAKEY, L. and HEATH, A. (1992). 'Differences between comprehensive schools: some preliminary findings'. In: REYNOLDS, D. and CUTTANCE, P. (Eds) *School Effectiveness: Research, Policy and Practice.* London: Cassell.

BROWN, S. (1995). 'School effectiveness research: the policy-makers' tool for school improvement?', *EERA Bulletin*, 1, 1, 6–15.

FITZ-GIBBON, C.T., TYMMS, P.B. and HAZELWOOD, R.D. (1989). 'Performance indicators and information systems'. In: REYNOLDS, D., CREEMERS, B. and PETERS, T. (Eds) *School Effectiveness and Improvement: Proceedings of the First International Congress.* Groningen: RION.

FULLAN, M. and HARGREAVES, A. (1992). *What's Worth Fighting for in Your School? Working Together for Improvement.* Buckingham: Open University Press.

FULLAN, M. with STIEGELBAUER, S. (1991). *The New Meaning of Educational Change.* London: Cassell.

GOLDSTEIN, H. (1995). *Multi-Level Statistical Models* (2nd Edn). London: Edward Arnold.

GOLDSTEIN, H. and HEALY, M. (1995). 'The graphical presentation of a set of means', *Journal of the Royal Statistical Society*, Series A, 158.

GRAY, J. and WILCOX, B. (1995). *'Good School, Bad School': Evaluating Performance and Encouraging Improvement.* Buckingham: Open University Press.

GRAY, J., JESSON, D., GOLDSTEIN, H., HEDGER, K. and RASBASH, J. (1995). 'A multi-level analysis of school improvement', *School Effectiveness and School Improvement*, 6, 2, 97–114.

GRAY, J., REYNOLDS, D., FITZ-GIBBON, C. and JESSON, D. (Eds)(1996). *Merging Traditions: The Future of Research on School Effectiveness and School Improvement.* London: Cassell.

HOPKINS, D., AINSCOW, M. and WEST, M. (1994). *School Improvement in an Era of Change.* London: Cassell.

OFSTED (1994). *Improving Schools.* London: HMSO.

REYNOLDS, D. (1996). 'Turning around ineffective schools: some evidence and some speculations'. In: GRAY, J. *et al.* (Eds) *Merging Traditions: The Future of Research on School Effectiveness and School Improvement.* London: Cassell.

SCHOOL CURRICULUM AND ASSESSMENT AUTHORITY (1994). *Value Added Performance Indicators in Schools.* London: SCAA.

TEDDLIE, C. and STRINGFIELD, S. (1993). *Schools Make a Difference: Lessons Learned from a Ten-Year Study of School Effects.* New York: Teachers College Press.

## CORRESPONDENCE

Professor John Gray, Homerton College, Cambridge, CB2 2PH.