

## *Of Trends and Trajectories: searching for patterns in school improvement*

**JOHN GRAY**, *Faculty of Education, University of Cambridge*  
**HARVEY GOLDSTEIN**, *Institute of Education, University of London*  
**SALLY THOMAS**, *Graduate School of Education, University of Bristol*

**ABSTRACT** *The article builds on earlier research suggesting that it is fairly difficult to extrapolate current trends in school performance into the future (see Gray, Goldstein & Thomas, British Educational Research Journal, 27, 2001). It explores two areas of research on school improvement which have, to date, received relatively little attention—trends in performance over time and the incidence of time-lagged phenomena. In responding to a critique by Pugh and Mangan (2002), it suggests that, whilst of interest, time-trend analyses favoured by economists may be less useful than other forms of analysis which search for consistent patterns over shorter periods of time. Some strategies for thinking about their identification are put forward. The implications of these approaches for target-setting and related exercises, which are premised on linear assumptions about school improvement, are also discussed.*

### **Introduction**

Over the last few years the results of national assessments have continued to rise. This national trend has, somewhat unwisely in our view, been extrapolated to individual schools. In a climate of rising expectations there is a danger that every school will be expected to show year-on-year 'improvement' in its results. This research note explores some of the very limited evidence currently available on trends in schools' performance over time and sounds a cautionary note.

It is prompted by two separate but related developments. The first is the empirical discovery in our own research that, over the short run, next year's results at A level could not be predicted with any great confidence from trends in the same institutions over the previous three years (see Gray *et al.*, 2001). In a somewhat critical commentary on part of that research, Pugh and Mangan (2003) argue that even such trends as we discerned should not necessarily be extrapolated with any great confidence. The second development is an increasing awareness that some of the complexities of unravelling trends in schools' performance over time have yet to be properly addressed. This is

Received 6 August 2002; accepted 15 August 2002.

TABLE I. Theoretical incidence of ‘improvement trajectories’

Sequence number	After 1 year	After 2 years	After 3 years	After 4 years	After 5 years
1	U	UU	UUU	UUUU	UUUUU
2	S	US	SUU	SUUU	SUUUU
3	D	UD	USU	USUU	USUUU
4		SU	UUS	UUSU	UUSUU
5		SS	etc.	UUUS	UUUSU
		SD	to	etc.	UUUUS
		DU	DDS	to	etc.
		DS	DSD	DDSD	to
		DD	DDD	DDDD	DDDDD
Total combinations	3	9	27	81	243
‘upward’ trajectories		1	4	5	6

partly a matter of data sets on which empirical estimates might be based not yet being available. But it is also a question of expectations. Schools with ‘improvement trajectories’ in a *consistently* upward direction are probably much rarer than is commonly supposed. This issue can be explored theoretically and now demonstrated, to a limited extent, empirically.

### Patterns of Improvement?

We can define a school with a ‘consistently improving’ trajectory as one where the results go up year-on-year for a number of years. By way of example, schools’ changes in results could be coded into three equal-sized groupings, with the top third described as going ‘up’ (denoted by a U), the middle group as staying roughly the ‘same’ (S) and the third group as going ‘down’ (D).

Table I shows some of the resulting patterns and combinations which result from this simple approach. With a single year’s data the patterns are very straightforward; only one of the three (U) is, by definition, in the upward direction. With two years’ data, interpretation is still relatively simple—one of the nine possible patterns (UU) is consistently upwards. As data from further years are added, however, the incidence of *consistently* upwards trajectories diminishes rapidly. With four years’ data it is one out of 81 (UUUU) and with five years’ one out of 243 (UUUUU).

As Table I suggests, the insistence on ‘consistency’ in a trajectory imposes a particularly stringent set of assumptions on the data. If the assumption that there should be upward movement every year is relaxed to allow for a single ‘blip’ (an S in the sequence), then more patterns may potentially be identified as having an upward trajectory. But, nonetheless, their incidence is still rather low. With four years’ results, just 5 out of 81 could qualify (the first five listed in Table I) and with five years’, 6 out of 243 (the first six listed). The underlying assumption in the sequences laid out in Table I is that, at any particular point, the probabilities that next year’s results will go up, stay the same or go down are equal. In reality, the probabilities of this happening are not known. Our work has begun to provide some empirical estimates but other assumptions may be more appropriate. In particular, the possibility that there are *trends* over time in

schools' performance merits further investigation. The existence of such trends might suggest that there are more systematic underlying factors driving the changes.

### The Incidence of Trends

Time-series analyses are fairly common in economics but have rarely been used in educational research. To the best of our knowledge they have never been used in research on school effectiveness or school improvement. One of the main reasons for this is doubtless that the necessary databases, requiring long runs of results, have not been assembled. Our article (Gray *et al.*, 2001) was highly unusual for possessing data on as many as four cohorts. These enabled us to construct four separate value-added estimates. In due course, more extensive data sets may become available but typically time-series analyses have substantially more time-points (usually at least 10 or preferably more).

Pugh and Mangan (2003) point out that economists are typically interested in establishing whether trends are basically 'deterministic' or 'random'. In particular, they suggest that the data 'identified by the authors [Gray, Goldstein and Thomas] as school improvement trends may not be generated by the deterministic process they [we] assume. Instead, these upward and downward movements may reflect a purely stochastic (i.e. random) process or, at least, a process that allows for both stochastic and deterministic elements. In either case, once one treats the notion of trend as problematic, one has to allow for the possibility of a purely random element in school improvement that is greater than even the authors suggest'. They then go on to make some general points about fitting models to time-series data on examinations in order to predict future performance.

First, they describe a regression model of the form (using their notation):

$$x(t) = c + bt + u(t) \quad (1)$$

where the measurements are at the school level; in our article we obtain these by fitting a multilevel model. Their second model is a so-called 'random walk' model:

$$x(t) = x(t - 1) + v(t). \quad (2)$$

The random walk model (and more general 'stationary' models such as autoregressive models) assume that any underlying trends have been removed so that the correlation between measurements  $t$  time units apart depends on the assumed random structure, such as in (2) above. Knowledge of this correlation structure then allows a prediction from an initial value.

In our article we report the empirical (residual) correlation matrix across the four years of the study. Our aim was simply to use this in order to establish the efficiency with which predictions of subsequent school performance can be made from earlier performance data. Since we have extensive data for each year, we have accurate empirical estimates of this residual correlation matrix and so have no need to make any particular assumptions about the forms of any underlying time series model.

There was no intention on our part of extrapolating linear or non-linear trends, as in equation (1), beyond the observed time-scale so that Pugh's distinction is irrelevant to our main purpose. The linear regression model that we fitted to the school level effects for the first three years is simply a convenient summary of the trend in order to compare the predicted with the observed value. The point of the analysis, which compared summary trends for the four years across schools, was to demonstrate that such comparisons could not be used reliably to rank schools in terms of 'improvement'.

### The Primitive State of Knowledge about Patterns and Trends

There is growing worldwide interest amongst policy-makers in judging the performance of schools based on what Linn and Haug describe as ‘changes in the achievement of successive groups of students’ (2002, p. 29). Their research, which is based on much cruder measures of performance than the value-added measures we have developed, suggests that there is ‘a great deal of volatility’ in the year-on-year changes. They point out that changes in performance from one year to the next are driven by a variety of factors. Some of these are ‘real’ and genuinely reflect schools’ improvement efforts. Others may stem from such things as measurement error and changes in the student body from one year to the next, not all of which will necessarily be captured in even the more sophisticated analyses. Still others, they argue, may stem from so-called ‘non-persistent’ factors such as ‘changes in the teaching staff’. ‘These other factors’, they conclude, ‘make up a substantial part of the overall variability in school change scores’ (2002, p. 34).

Thus far their comments seem unexceptional. It is the conclusions which Linn and Haug then draw that pose a challenge for policy-makers and researchers alike. They state:

it should not be surprising therefore that schools identified in one change cycle as outstanding for achieving a large change in achievement are unlikely to repeat that performance in the next cycle. The converse is also true. Thus, schools that are identified to need assistance in one cycle because of falling short of their change target, or even showing a decline, are unlikely to fall in that category in the next change cycle. (p. 36)

They then go on to sound a warning about the over-hasty interpretation of trends in performance. They remark:

A consequence of this random fluctuation from one change cycle to the next is that the actions taken to assist schools in the latter category (falling short) may appear to be more effective than they actually are. Moreover, it is likely to be a mistake to assume that the practices of the schools recognised as outstanding are ones that should be adopted by other schools. (p. 36)

Our own research would tend to support the view that changes in schools’ performance over time are a good deal more modest and variable than is sometimes supposed (Gray *et al.*, 1996, 2001; Thomas, 2001). Only a small minority of institutions seems to improve consistently over time; the greater majority has ups and downs. Furthermore, where upward movement does occur, it is of relatively short duration. Three years seems to represent a ‘good run’; four or more years currently seems exceptional. Three years is, of course, the minimum needed to establish a *trend*. Ironically, if our rather tentative conclusions come to be replicated on other datasets, it would also seem to represent the *duration* of such trends. School ‘improvement’ would appear to be characterised by bursts of activity over relatively short periods of time.

The limited amount of work that has been undertaken to date on trends in schools’ performance over time has tended to focus on the identification of *linear* trends. Such trends may be easy to interpret but it is an empirical question how frequently they occur. If it turns out that the ‘bursts’ interpretation is valid, then other ‘stepped’ patterns may be just as (or equally) frequent. One example of such a pattern would be a run of changes which went SUUUS; another would be UUSSU. Current techniques, typically employed by economists, for determining whether there is an underlying structure in a run of data

TABLE II. Patterns and sequences in research on school improvement

	2000	2001	2002	2003	2004	2005	2006
<i>Outcome Measure</i>							
Pupils' examination results	S	U	D	U	U	U	S
Pupils' attendance	S	S	U	S	S	U	U
<i>Changes in schools' policies and practices</i>							
Policies on attendance		Change (1)			Change (2)		
Leadership	Change				Change		
Approaches to teaching		Change (1)			Change (2)		

Notes: S = 'same' as previous year; U = 'up' on previous year; D = 'down' on previous year. In the lower half of the table the word 'Change' indicates that there was a change in some aspect of the school's policies or approaches during the year.

may not be up to the task of identifying such patterns. More sensitive techniques may need to be developed.

### The Search for Time-Lagged Effects in School Improvement

The availability of data on trends over time will also permit more sophisticated explanations to emerge with respect to factors affecting school improvement. Most research on school effectiveness and improvement has hitherto been conducted using rather short-term perspectives and time-frames. More long-term data would enable us, for example, to look at the effects of changes in schools' policies as they unfold over longer periods than are usually considered. There again, changes in leadership have often been viewed as crucial in the school improvement literature. What has not been properly recognised, perhaps, has been the methodological 'bias' implicit in much time-constrained research, which has tended to favour (albeit unintentionally) the identification of high-impact, short-term leadership styles as potentially more effective—possibly at the expense of their longer-term consequences (cf. Gray, 2000). Table II looks forward to the period when more data might become available and illustrates some of the possibilities 'time-lagged' analyses might reveal. The notation is the same as that employed in Table I earlier.

Hitherto, almost all studies of school effectiveness and improvement of a quantitative kind have operated within a limited time-frame. They have basically attempted to explain changes and improvements across two or (occasionally) three of the columns in Table II. By way of example, a change of headteacher in 2000 (identified in the table as a change in 'leadership') might typically be linked with a rise in examination results in 2001 (indicated by the letter U in the first row). There is certainly some empirical evidence that newly arrived heads behave in this way; they look for so-called 'quick wins'. By focusing their energies on Years 10 and 11 they may well be able to effect some immediate changes. In this example, however, the rise is followed by a fall the following year.

What is of greater interest in this example is the sequence UUU, which occurred between 2004 and 2006. In the hypothetical example described in Table II, the change of leadership represents the start of a longer stream of events. These first brought about

changes to the school's attendance policies and teaching which, in turn eventually fed through to higher levels of pupil performance.

### Conclusions

Interest in how schools perform over time is growing but empirical evidence is currently in short supply. This article is intended, therefore, to anticipate some of the issues which will need to be considered when longer runs of data become available. As Pugh and Mangan suggest, insights and analytical procedures taken from economics may be relevant, but whether these will be sufficient seems doubtful. In the meantime, a degree of caution would seem appropriate. More needs to be established about patterns of performance with particular reference to their length and duration. In particular, the assumption that there are *consistent* patterns may prove unjustified. Given our current understanding of the evidence, it seems more probable that school improvement comes in relatively short 'bursts'. It follows that attempts to impose a predetermined (and upward) 'trajectory' onto schools' performances (such as are implied in target-setting and comparable exercises) are probably premature.

*Correspondence:* Professor John Gray, Faculty of Education, University of Cambridge, Cambridge CB2 2PH, UK; e-mail: jmg1004@cam.ac.uk

### REFERENCES

- GRAY, J. (2000) *Causing Concern but Improving: a review of schools' experiences* (London, Department for Education and Skills).
- GRAY, J., GOLDSTEIN, H. & JESSON, D. (1996) Changes and improvements in schools' effectiveness: trends over five years, *Research Papers in Education*, 11, pp. 35–51.
- GRAY, J., GOLDSTEIN, H. & THOMAS, S. (2001) Predicting the future: the role of past performance in determining trends in institutional effectiveness at A level, *British Educational Research Journal*, 27, pp. 391–405.
- LINN, R. & HAUG, C. (2002) Stability of school-building accountability scores and gains, *Educational Evaluation and Policy Analysis*, 24, pp. 29–36.
- PUGH, G. & MANGAN, J. (2003) What's in a trend? A Comment on Gray, Goldstein and Thomas (2001), *British Educational Research Journal*, 29, 77–82.
- THOMAS, S. (2001) Dimensions of secondary school effectiveness: comparisons across regions, *School Effectiveness and School Improvement*, 12 (4).