

# School Effectiveness and 'Value-added' Analysis

Harvey Goldstein & Sally Thomas

Harvey Goldstein and Sally Thomas are associate directors of the International School Effectiveness and Improvement Centre at the Institute of Education, University of London.

In November 1992 schools' 'raw' examination results were published for the first time by the Government and reproduced in most national newspapers, fulfilling a pledge made in the Parents' Charter to provide more information about schools. At the same time, these league tables were widely criticised as unfair and misleading due to the differing intake of many schools. There has for some time been a considerable weight of research evidence which shows that by far the best predictor of student GCSE and A-level exam achievement is the achievement of the student on entry to their secondary school. Schools which have high average exam results tend to be those whose students have high achievement when they start. Ranking schools on the basis of their exam results then, in large measure, merely reflects attainment at entry. This did not prevent the Government, when it introduced the Parent's Charter, from claiming that the exam result league tables would result in fair comparisons of the effectiveness of school and the quality of their education.

Several years on, the idea of 'value-added' in educational terms has been accepted by a large section of society: parents, teachers, researchers, LEA officers, journalists, employers, and even government policy makers. In comparison to 'raw' examination results, 'value-added' indicators - quantitative measures of the educational progress pupils make while at school - can in theory provide accurate information for making choices about individual schools for individual pupils or, more importantly, when we wish to evaluate schools for the purpose of improving educational standards.

What is value-added analysis? We shall illustrate our discussion using results from an analysis of information collected about A-level results in Summer 1993. This was obtained as a result of a collaboration between the Institute of Education, University of London, and *The Guardian* newspaper and published by the latter on November 30, 1993, to provide a rational alternative to the Government's crude league tables of examination results. A full discussion is given by Goldstein & Thomas (1995).

Figure 1 shows hypothetical relationships between student GCSE examination scores and A-level scores for two schools. The vertical axis represents the predicted A-level score in each school for students with different GCSE scores. With increasing GCSE score so the predicted A-level score increases, but note that whereas for low GCSE scores we have one school with a lower predicted A-level score at each GCSE score, this school obtains higher predicted A-level scores for those students who obtain high GCSE scores. This 'differential effectiveness' for students with differing levels of prior achievement is commonly found in school effectiveness studies. It may also extend

to other student characteristics. For example, some institutions may be relatively more effective for girls than boys, some for particular ethnic minority groups, etc. Furthermore, individual departments within schools appear to be differentially effective: Goldstein et al (1993) found in an analysis of the GCSE exam results for Inner London schools, that there was very little relationship between the effectiveness, or value-added, scores for Mathematics and English. Similar results were found in a more recent study across six different subject outcomes (Thomas et al, 1995). Likewise, during the course of schooling students will encounter different teachers who will vary in their 'effectiveness', and differentially so for different kinds of students. Research has also indicated that, for some schools, results can vary substantially over as short a period as 3 years (Thomas et al, 1995).

We see, therefore, that the reality in schools is complex and cannot be summarised by a single measure, whether value-added or not, without distorting what a school is achieving. It is, of course, better to present a value-added estimate for an institution rather than a 'raw' average; but in addition to the limitations we have already mentioned,

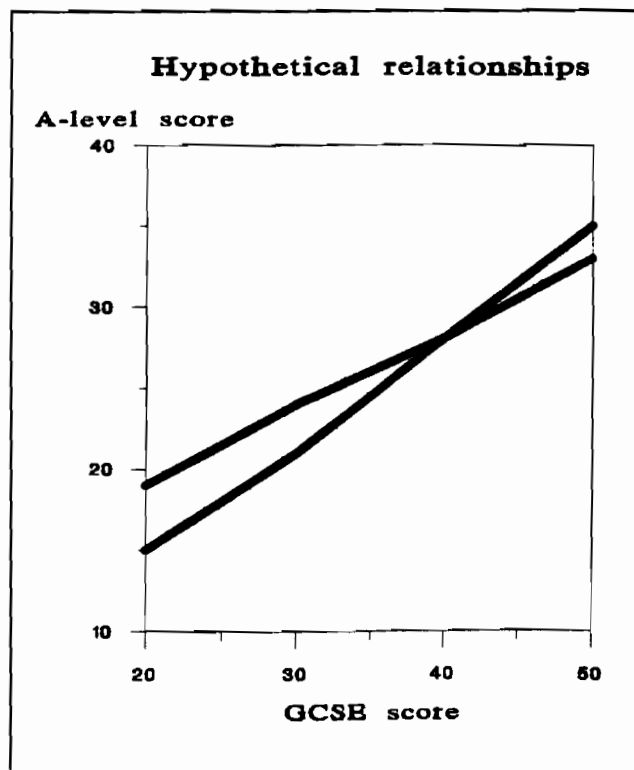


Figure 1

there are two further ones which severely limit the use of value-added (and of course raw) estimates for purposes of accountability. By this we mean essentially placing institutions in some kind of rank order so that judgements between any pair of them can be made.

First of all, in the case of A-level results, suppose we have available, in late 1994, value-added estimates for Summer 1994 examinations. These results refer to a cohort of students who began their A-level courses in Autumn 1992. If we wish to base decisions on a choice of institution for students starting in 1995 then the results are essentially three years out of date. In the meantime those institutions and departments within them may have changed, become more or less effective and so the historical information will be limited by this inevitable time delay. This problem is more severe if we wish to judge secondary schools using value-added measures calculated between intake and outcome (GCSE) five years later since results will then be six years out of date!

The second problem is this. While we can study the factors associated with student performance and come to conclusions about which of them appear to be associated with 'success', yet it seems to be very difficult to identify precisely *which* schools are doing well or badly. We can sometimes identify a few 'outlying' or 'extreme' schools. There are limited numbers of students in any one school year, or even over, say, a three year period and using factors such as intake achievement and social background we can only explain a proportion (typically about a half) of the variation in outcome scores. Where we are able to obtain value-added estimates for each school they are mostly too imprecise to provide a fine separation of institutions. Research on GCSE results (Goldstein et al, 1993, Goldstein & Healy, 1995) suggests that about two thirds of all the value-added comparisons between pairs of schools are of this kind; the institutions cannot be ranked. Figure 2 illustrates this point. Schools are to be judged as statistically significantly different (at the 5% level) only when the 'bars' for a pair of schools do not overlap. It is clear that we can only sensibly make use of the information for the schools at the extremes; even then our inferences must remain tentative.

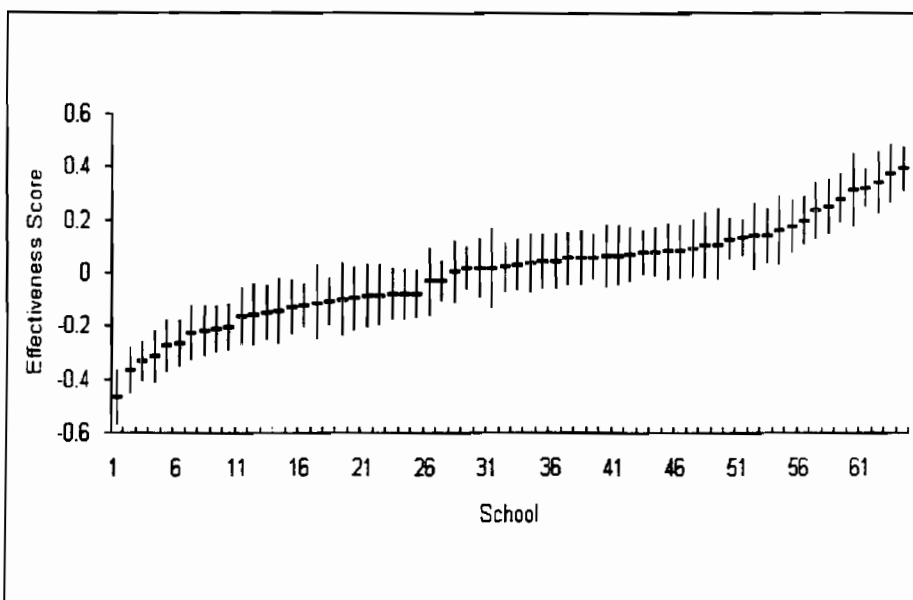


Figure 2

A kind of 'uncertainty principle' operates in which we can make useful statements about why schools differ, without necessarily being able to pinpoint precisely which schools are different. In other words, research into school effectiveness is a useful activity in our attempts to obtain knowledge about the process of education, but a very poor tool for holding schools to account.

The implications of all this for current debates is fairly clear. The use of test or exam scores to judge schools publicly is insupportable. Information provided by 'raw' exam or test score league tables is misleading and potentially educationally damaging. Value-added league tables eliminate one source of error but retain others that also make them insupportable. Used appropriately, however, by LEAs or other responsible agencies as a confidential screening instrument, a value-added analysis can provide a source of relevant information that can be used along with other evidence to evaluate the relative performances of schools for different groups of pupils (Thomas & Mortimore, 1994). Like all screening instruments, it can only provide a first indication of which schools may be the very high or low performing ones and should not be used to provide definitive judgements or firm 'labels'. In addition, the evidence points strongly to an approach where value-added information should be accumulated over several years, using a wide range of outcomes to reflect the complexity of schooling.

In our view, the use of value-added or raw results as a major component in identifying schools for 'improvement' interventions is problematic since it is predicated upon the assumption that they can identify 'failing' schools with some precision. Even where value-added results are used privately by individual schools, care is needed to avoid assuming an accuracy which the data do not support. This is especially true where estimates are obtained for individual departments where there may be very few students taking a subject exam, with a consequent large measure of uncertainty attached to any rank ordering. If a school is fully aware of this uncertainty however, a value-added estimate may be useful as one piece of information among others about its performance.

The purpose of this article has been to point out the limitations of current attempts to provide public accountability measures using examination and test score results. There remains the question of what to substitute in their place; while any answer to that question requires intensive study and debate it appears to us that a fruitful direction is one which looks for measures of process related to teaching and learning. To some extent this is what is done during OFSTED inspections, but much more is required by way of a thorough study and evaluation of alternative approaches to measuring process in a standardised way. Of course, there is no guarantee that even if it becomes possible reliably to determine where poor teaching occurs, that this will also tell us where poor outcomes are to be found. We

understand little about the subtle interaction between student progress and the process of schooling. This is an important area for further research.

#### References

- Goldstein, H., Rasbash, J., Yang, M., Woodhouse, G., Pan, H., Nuttall, D.L. & Thomas, S. (1993) A multilevel analysis of school examination results, *Oxford Review of Education*, 19, pp. 425-433.
- Goldstein, H. & Healy, M.J.R. (1995) The graphical presentation of a collection of means, *Journal of the Royal Statistical Society, A*, 158, pp. 175-177.
- Goldstein, H. & Thomas, S. (1995) Using examination results as indicators of school and college performance, *Journal of the Royal Statistical Society, A*, 158 (forthcoming).
- Thomas, S. & Mortimore, P. (1994) Report on value-added analysis of 1993 examination results in Lancashire, *Research Papers in Education* (forthcoming).
- Thomas, S., Salmons, P. & Mortimore P. (1995) Stability and consistency in secondary school effects on students' outcomes over 3 years. Paper presented at ICSEI, January, Leeuwarden, Netherlands.

# Towards a Learning Community

**Pat Collarbone & Maggie Farrar**

The Headteacher, Pat Collarbone, and her Deputy, Maggie Farrar, have collaborated in writing about significant developments within Haggerston School, an inner city 11-16 comprehensive school committed to raising the achievement of staff and students together.

A recent job specification for a post in Haggerston school included as part of the selection criteria "Must truly wish to take part in and belong to a community of learners". What does this mean and why was it important enough to be included as part of the selection process for a deputy head?

We were first introduced to the idea of a school as a community of learners through the ideas of Professor Roland Barth:

*The school as a community of learners is a place where all participants – teachers, principals, parents and students – engage in learning and teaching. School is not a place for important people who do not need to learn and unimportant people who do. Instead school is a place where students discover, and adults rediscover, the joys, the difficulties, and the satisfactions of learning.[1]*

These ideas came to the school at a time of rapid and turbulent change two years ago. As a school in inner London the years following the Education Reform Act in 1988 had brought us a diet of growing autonomy, LMS and school development planning. This was accompanied by a new vocabulary of action planning, aims, objectives and success criteria but without the sustenance of the ILEA. We were aware that in order for a school to cope with this amount of change and autonomy we needed to work with teachers to ensure that they:

*engage in frequent, continuous and increasingly concrete and precise talk about teaching practice. By such talk, teachers build up a shared language adequate to the complexity of teaching, capable of distinguishing one practice and its virtue from another.[2]*

In the words of Caldwell & Spinks we were a self managing school:

*one that for which there has been significant and consistent decentralisation to the school level of authority to make decisions related to the allocation of resources.[3]*

Drawn by their lists of 'characteristics of effective

schools', and also the 'collaborative school management cycle' we sought to make sense of the post ERA world. We were also concerned not to be drawn by the list logic and recipe book approach to school improvement much in vogue at the time. Most importantly, we were concerned that the staff in the school, whilst buffeted by change, were still striving to teach in a creative and accessible way while being hampered by a content laden curriculum. We wanted to enable them to feel energised, rather than overwhelmed by, the scale and speed of the changes they were faced with. We wanted to find a way of reaching the position described by Hargreaves & Hopkins when they say:

*A school should reach a point where change is not something extra or unusual but a task with which it can cope comfortably because innovation and change have become part of management.[4]*

Our reading of Hargreaves & Fullan's *What's Worth Fighting for in Your Schools* led us to the notion of 'stuck' and 'improving' schools. The description of the 'moving school' seemed to have direct parallels with the notion of a learning community:

*Teachers work together more. Most teachers, even the most experienced, believed that teaching was inherently difficult. They believed teachers never stopped learning to teach. Since most teachers acknowledged that teaching was difficult, almost everyone recognised that they needed help. Giving and receiving help was part of the common quest for continuous improvement.[5]*

This struck a chord with us. We wanted to be moving. As our discussions on a learning community progressed we began to look at the seeds that had already been sown. Did we have instances in the school where the dual culture of the learned and the learners had already begun to be broken down?

For some time we had been engaged with our local higher education institutions in partnership work in the training of initial teachers. It was clear to us that if young teachers were going to become reflective and critical practitioners then they had to be trained in an environment