



Curriculum, Evaluation and Management Centre

PUBLICATION NO.24

PERFORMANCE INDICATORS AND INFORMATION SYSTEMS

C.T. FITZ-GIBBON
P.B. TYMMS
R.D. HAZELWOOD

School Effectiveness and Improvement
1989

PERFORMANCE INDICATORS AND INFORMATION SYSTEMS

C.T. FITZ-GIBBON, P.B. TYMMS AND R.D. HAZELWOOD

1. A Three Step Approach for Research on School Effectiveness

An attractive approach to improving school effectiveness involves three steps:

1. locating effective schools;
2. finding the educational processes which make them effective; and
3. transplanting these processes to other schools.

This strategy is clear and appealing. It appears to emphasise good practice and to be efficient in that only effective schools need to be studied in depth. Unfortunately there are several reasons why this three-step strategy may be neither viable nor useful.

Consider the first step: locating effective schools. An immediate question is, "Effective in what way?" There are many goals for schools: smooth functioning, high attendance, examination successes, satisfied pupils and parents, desirable attitudes, absence of racism, curriculum breadth and balance, etc. It cannot be assumed that a school doing well with respect to one goal is necessarily doing well with respect to all the other goals. Indeed, there may sometimes be trade-offs to be made, consciously or unconsciously. For example, some schools may choose to spend money on residential experiences for all pupils rather than smaller classes for academic subjects. Are they emphasising other goals than academic ones?

Each of the goals which schools choose to address will have to be considered separately for otherwise impossible problems arise in trying to weight the importance of the various goals. And of course it is necessary to consider the attainment of each of the goals in the light of the situation in which the school functions. In particular, account must be taken of the kinds of pupils enrolled.

Suppose, however, that the problem of goal definition is simplified by agreeing one major goal for effectiveness. Such a goal might be that a school should obtain for its pupils good examination results. (But what are "good" results? Appropriate to the pupils' abilities or better than you would expect knowing the pupils' abilities? We shall leave this question aside for now and simply agree that the outcome of examination results shall be the variable of interest in assessing effectiveness.) Suppose, then, that life has been easy and not only has a dependent variable been agreed but also a relevant co-variate to create statistical controls for intake differences. We can even agree to leave to statisticians the problems of analysis methods (Aitkin and Longford, 1986; Goldstein, 1988; Cuttance, 1985; Raudenbusch and Bryk, 1986), but we still have problems. Schools which show up as effective one year may not be noticeably

effective another year: i.e. effectiveness measures may not be stable. If the rank ordering of schools changes from year to year, this must leave us a little uneasy about designating a school as "effective" and worth copying. Next year it might not do so well. Furthermore the various methods of statistical analysis may also affect the rank ordering (Goldstein and Cuttance, 1988).

Figure 1 shows effectiveness measures for six schools across five years for two Advanced level General Certificate of Education subjects: English and mathematics. It can be seen that schools effective on one subject were not necessarily the same as the schools effective in the other subject. Perhaps then it would be wise to consider examination results subject by subject rather than summed within schools to give an index. This would imply a shift from the school as a unit of analysis to the subject departments within schools. More data will be needed before this can be a firm conclusion but it seems quite a feasible proposition that departments within schools may have greater effects on examination results than the school itself. Departments, moreover, are perhaps more manageable units, and therefore more accountable, than whole schools. Ideally a multi-level model will consider pupils within departments within schools and compute effects at each level.

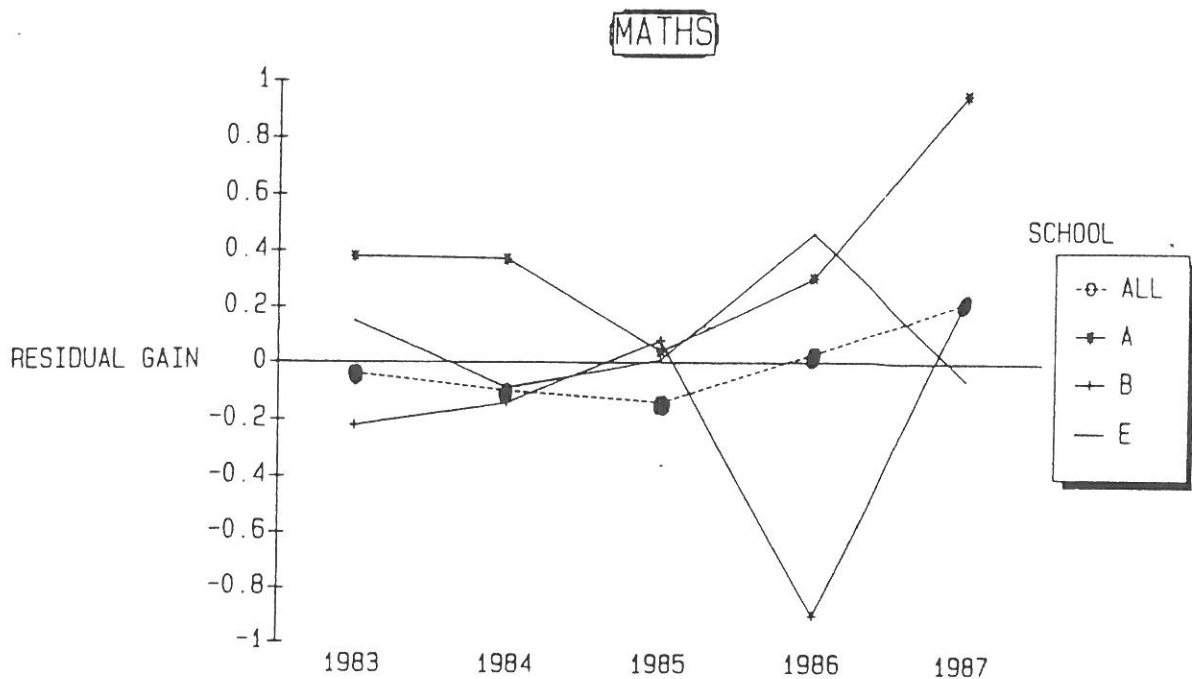


FIGURE 1A

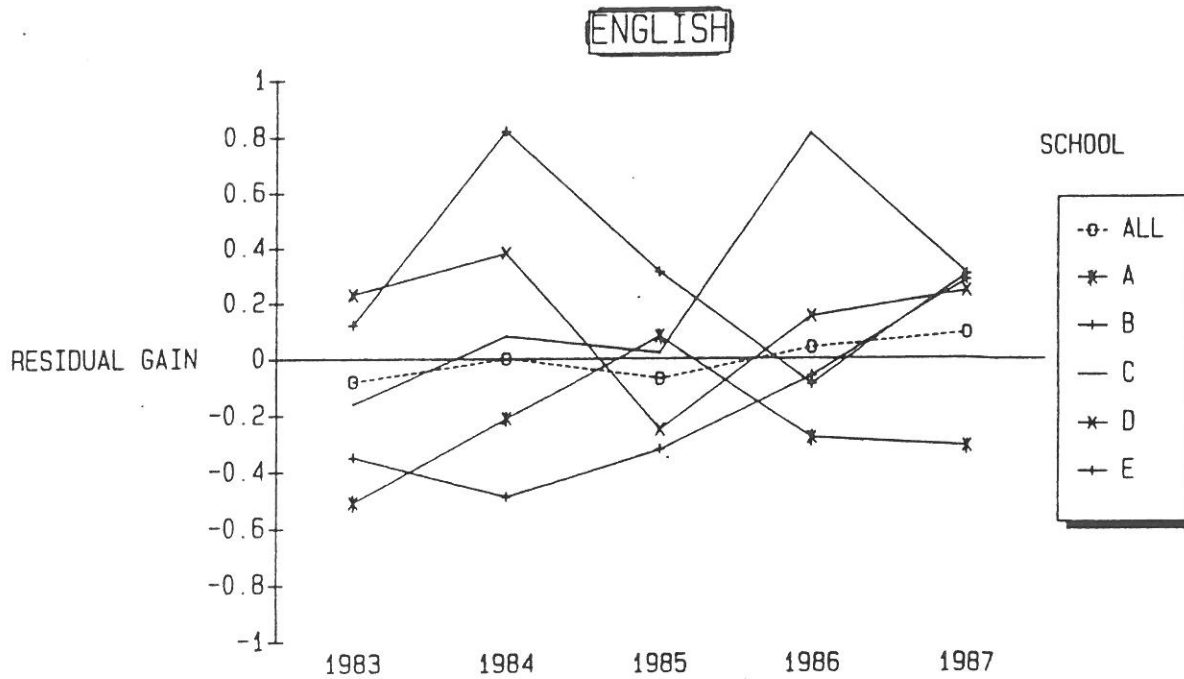


FIGURE 1B

Can we locate effective departments, then, rather than schools? Once again we confront the instability of the data. The knowledge as to which departments were the more effective in any particular year only becomes available AFTER the teaching has occurred. To locate the teaching processes which made the department effective in that particular year, the investigators would have had to have documented processes in ALL schools being investigated in order to be able to describe the processes in the most effective schools retrospectively. So the efficiency which is one appeal of the three step model breaks down in the face of year-by-year instability.

The instability may not be surprising since schools are far from static organisations and there are many possible influences on examination results: Who teaches the subject; the teaching methods used this year; the examination board and syllabus adopted; the textbooks; which classes had supply teachers; which groups had classes Friday afternoons; the ethos developed for a class ... these and many more factors which might influence outcomes may all vary from year to year. Can the school effectiveness researcher collect and weight all this information? If he or she did, other factors might be remembered later as explanatory (e.g. "in that class they all hired private tutors"). Clearly we could wait years for this researcher-as-interpreter model to pay dividends.

The notion that having found an effective school (or department) one will then be able to locate the processes which made it effective, is highly optimistic. Need it be said again that correlation is not causation? If it were so simple to establish cause and effect in Education we should not be searching for effective processes, we would be implementing them. Indeed, discrete, transferable processes may not exist. It may be that complex configurations are needed for effectiveness. The best we can hope for from step 2, as researchers looking for generalisations or principles, is the generation of hypotheses which would need to be tested experimentally. The need for basic research is evident. However, we concur with Glass (1979):

I am definitely not advancing a nihilistic position ... complexity and incomplete knowledge do not mean that there is no hope. Hope may lie elsewhere than where we are customarily looking for it when we seek to illuminate educational policy by the light of educational research.

As we shall shortly discuss, it is to the practitioner that we must turn for effective use of the information on a year to year basis.

Step 3 has its own difficulties. Even if we found some particular processes used in some consistently effective schools (or departments), and even if we felt there were strong reasons for believing that these processes directly affected outcomes in a positive manner, there is no guarantee that the processes would prove to be transplantable to a different terrain. And what level of resources would be needed to give such an effort of transplantation a fair chance to succeed? And if the processes identified by researchers transplanted successfully but the predicted improvements in "effectiveness" did not follow, how would this affect the reception of subsequent recommendations from researchers?

All in all, we no longer find the three-step strategy appealing. So what strategy should be offered to improve schools?

2. An Alternative Strategy

An alternative strategy is to implement on-going "Quality Control" procedures: information systems which feed valid and useful information to all levels of the educational system, particularly information on important outcomes

such as examination success. It is then the prerogative and the responsibility of the various management levels in the system to interpret and use the data in the light of their in-depth knowledge of the situations pertaining in the units which they manage. In the type of information system we envisage, the users of the information will also assist in revisions of the data collection instruments and in some of the research arising out of hypotheses generated by the quality control data.

How does this new 'customer-led' approach differ from the three-step approach just considered? In the first instance the data analyst has moved from the role of expert and interpreter to the role of partner in a continuous cycle of information - action information. This move does not negate the problem associated with school effectiveness studies outlined in the previous section but aims to treat the findings as an interesting addition to the range of other data available to the manager on the ground. The move also helps the researcher because the data is easier to collect in an atmosphere of co-operation and may, in certain instances, be more accurate. Furthermore the possibility of monitoring interventions in a systematic way becomes possible. The importance of this cannot be overemphasised since there is the ever present danger that an association of factors will be assumed to imply a causal relationship whereas it is only by disturbing a system and observing the effect that one can know what is likely to happen if the system is disturbed. Such disturbances may arise in the natural course of events but they may be deliberately engineered and could be suggested by the researcher drawing on stable findings from the literature and from the data as it accumulates.

The open relationship between school and researcher should also help to investigate the problem of whether or not schools can affect the indicators for which they may one day be held responsible. This may in turn help to guard against the unfair use of performance indicators. For example, can schools be held responsible for delinquency or truancy rates? The fact such variables cluster does not mean schools are responsible for truancy or delinquency any more than they are responsible for clusters of leukemia. Again, correlation is not causation.

Nevertheless, data might be collected on such indicators, but only so that schools could TRY to affect them and see if, over the years, what they were trying appeared to work.

The example just given raises another issue which was highlighted by Harrison (1988). Some variables may be too difficult to include as Performance Indicators. Delinquency is notoriously difficult to measure but truancy could be measured economically in schools. Schools could be expected to monitor truancy more vigorously if the data were kept confidential and not reported to the LEA.

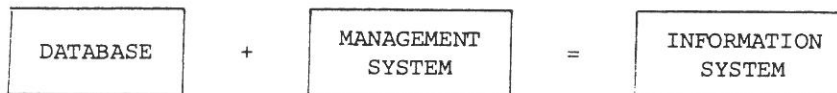
The use of an information system for quality control has always been recognised as desirable but the advent of computing power at reasonable costs makes it finally feasible. In the 1979 article already cited in which Gene Glass suggested that the education system was an unpredictable system, he outlined how one deals with unpredictable systems (drawing an analogy with how

we deal with outbreaks of fire... an unpredictable system). The characteristics he listed might well be appropriate to an information system for quality control:

Some general principles are recognisable in how we attempt to deal with unpredictable systems. Such systems must be monitored diligently; the actors within them must remain versatile and flexible, and the services must be highly decentralised. Persons must command options instead of eternal truths. In education, this style of coping would be the very antithesis of the style of top-down planning and policy that we researchers imagine ourselves discovering. (Glass, 1979, p.14)

The creation of the database for a quality control system is equivalent to the "diligent monitoring". The management teams in schools must receive the information and respond in ways which are "versatile and flexible". The data must be treated as just one more source of inspiration for "options", rather than as some firm finding, let alone "eternal truth".

This requires that a database is in operation alongside a management system which allows confidentiality, which feeds data to the level at which action can be taken and which constantly consults participants to gain from their intimate knowledge in order to revise and improve the information system:



Is a quality control information system feasible? Would it be too costly, or too disruptive? We certainly believe that systems can be developed immediately because we have been operating what might be seen as prototypes of such systems.

3. Examples of Information Systems

We conclude this paper with accounts of some of the performance indicators we have examined and some hypotheses generated in the course of operating two information systems.

3.1 The Combse A-level Study

In 1983 comprehensive schools in two LEAs were invited to join a study entitled CONFIDENTIAL, MEASUREMENT-BASED, SELF EVALUATION (COMBSE). What was on offer was the chance to participate in an information system providing each school with fair performance indicators for A-level provision. (A-levels ... Advanced level General Certificate of Education examinations are taken by 17

FIGURE 2

CONTENTS

1. INTRODUCTION	1
1.1 Participating schools and colleges	1
1.2 Data collection	1
1.2.1 Revision of the pupil questionnaire	1
1.2.2 Administration of the pupil questionnaire	1
1.2.3 Aptitude testing	2
1.3 Organisation of this report	2
2. CHARACTERISTICS OF THE A-LEVEL CANDIDATES	4
2.1 Prior Achievement: O-levels	4
2.2 Socio-economic status (home background)	7
2.3 Aspiration levels	10
2.4 Ability measures	14
3. SCHOOL PROCESSES	16
3.1 Class size - the size of teaching groups	16
3.2 Differences in the "pulling power" of the courses offered	16
3.3 Persons to whom the pupils can turn for help	16
3.4 Homework and other aspects of the use of time	19
4. ATTITUDES	25
4.1 Attitudes to the school	25
4.2 Attitudes to the subject	27
4.3 Extra-curricular activities	30
4.4 Classroom processes	30
5. RESULTS: A-LEVELS	35
5.1 A-level Grades obtained at the various schools	35
5.2 RESIDUAL SCORES	37
5.3 Teacher Expectations, Predictions	40
5.4 Other correlates of residuals	44
APPENDIX A : REPORTS PRODUCED IN THE COMBSE STUDY	47
APPENDIX B : GUIDE TO READING TABLES	48
APPENDIX C : THE SCALES USED	49

or 18 year olds after two year pre-university courses in schools, similar to US Advanced Placement examinations.) Data would be reported under code names for each school so that schools could compare their inputs and outputs with other schools while confidentiality was maintained between schools. Decisions to join were taken by school personnel and LEA officials were not involved after they had given permission for schools to be approached. The LEA did not receive the yearly data. About fifty per cent of the comprehensive schools approached joined the study as a result of simply one letter of invitation. This made a sample of 12 comprehensive schools. The researcher had no prior contacts with the participants and there were no efforts initially to increase the sample beyond those joining at the first invitation (partly because the study was considered a pilot and no funds had been applied for). This initial interest on the part of individual schools combined with their continued participation, with no withdrawals, suggests a genuine interest in the kind of data on offer: comparative, fair performance indicators. We would suggest that it is on this genuine professional interest in self-evaluation that an information system must be built.

The study was confined, until 1988, to A-level mathematics and English results. Data were specially collected from each school by a questionnaire and an ability test. This data collection was highly standardised and conducted by the researchers. (The standardisation of the conditions for the data collection was vital otherwise there was the risk of test-administration effects being confounded with the school effects of interest.)

Schools in the study receive a yearly report containing some 25 tables. As can be seen from the Table of Contents reproduced from one such report (Figure 2), the information system deals with more than simply examination performance. School processes are assessed. At first the processes assessed were just class size, sources of help, the "pulling power" of mathematics versus English departments and the allocation and use of time, but more recently we have been asking about teaching methods in the different classes.

Input characteristics which are measured include home background, educational aspirations and ability (measured under standardised conditions in each school by the researchers). Output indicators include attitudes to the school and to the academic subject and a measure of participation in extra-curricular activities as well as examination results. Thus multiple goals are considered.

3.1.1 A few findings from the A-level information system

In the literature on school effects there are plenty of warnings that schools account for very little of the variance between pupils. Often the figure of 10 per cent is cited (e.g. Willms, 1987). In data from four years of the study (1983-1986), the variance associated with schools for four outcome measures: academic grades, attitudes to the subject, attitudes to the school and reported levels of effort showed proportions ranging from less than one per cent (attitudes to A-level English and level of Effort reported by pupils for A-level mathematics) to more than 10 per cent (attitudes to mathematics and attitude to the school). Thus it appeared that schools had

more influence on pupils' attitudes than on academic achievement. This must, however, be viewed as a very tentative conclusion, awaiting more data.

Some other findings are that there were substantial changes from year to year in the rank-ordering of schools. Schools getting good results in English were not necessarily the ones getting good results in mathematics so that the department seemed to be the desirable unit of analysis rather than the school.

Examination results corrected for ability were perhaps the "fair performance indicators" of greatest interest. These corrected examination results are known as 'residual gains' as they index the extent to which a pupil gained a higher or lower grade than would be expected on the basis of his or her ability, i.e. the score left over (residual) after accounting for ability. The relationship of these residual gains to such process variables as class size and the amount of pupil-talk in class generated some hypotheses such as "Small classes are effective in mathematics but bad news in English". The Rosenthal expectancy effect (Rosenthal and Jacobson, 1968) was present in reverse: teachers making the most pessimistic predictions had the better residual gains.

3.2 The Newcastle TVEI Evaluation Studies

The methodology of using specially collected data from each pupil and of having a current measure of ability was also applied to our evaluation of TVEI (McCabe, 1986). One of the virtues of having pupil-level data is that it is immediately useful for the generation of many kinds of information. For example, if pupil-level data were always available, much evaluation effort could be saved by simply examining the database to look for the effects of special innovations on the yearly data. In the present case, when embarking upon an evaluation of TVEI, we made sure to collect data not only on TVEI pupils and on what the NFER refers to as 'non-TVEI pupils' (pupils in TVEI schools but not taking the TVEI course) but also on pupils in schools which were at the time not involved in TVEI at all. Thus a database was constructed for the evaluation of the Technical and Vocational Education Initiative consisting of over 2,000 pupils in the second of four yearly cohorts in "the TVEI pilot". Their examination results were added to the database in 1988 and proved somewhat surprising. (Although we must note that it is surprising how unsurprising data becomes once it is known.)

Looking at residual gains as "fair performance indicators" the alarming finding was that non-TVEI pupils obtained better results than TVEI pupils (Fitz-Gibbon, Hazelwood, Tymms and McCabe, 1988).

Why should this be alarming? The initiative represented the spending of almost unprecedented amounts of money. Did all this money and all the enthusiasm and effort that went into TVEI, depress examination results? If you believe in a simple cause-and-effect interpretation of performance indicators you would have to believe that TVEI represented negative effectiveness at very high cost. TVEI is now being extended to all schools (although without equivalent amounts of money) without any delay to examine the results of the

remaining years of the 'pilot'. The same model of money-on-offer-for-curriculum-changes (curriculum changes such as introducing experienced-based learning and enterprise projects relevant to the world of work) is now being implemented in the University sector.

However, the caveats that we have already expressed apply here. Schools have many different goals and so did TVEI. (Although an increase in the number of qualifications pupils obtained was an explicit goal of the initiative.) How are the various goals to be weighted in importance? Schools alone know the details of how TVEI was implemented and whilst the results are so consistent as to suggest a definite TVEI-effect of about -0.25 (using effect sizes as per Hedges and Olkin, 1985; Glass, McGaw and Smith, 1981) this might have been due to differences in the pupils opting for TVEI. We have, of course, applied all the corrections that we could (for ability, attitudes, gender, etc.) but no statistical correction can make a quasi-experiment into a true experiment with strong inferences. The findings from these information systems MUST be viewed tentatively, as information to take into consideration. That is how we see the Performance Indicators for TVEI: as an interim report which raises questions on a national level because of the very consistency of the data. Naturally, however, the local schools need also to look at their practices and watch for next year's indicators.

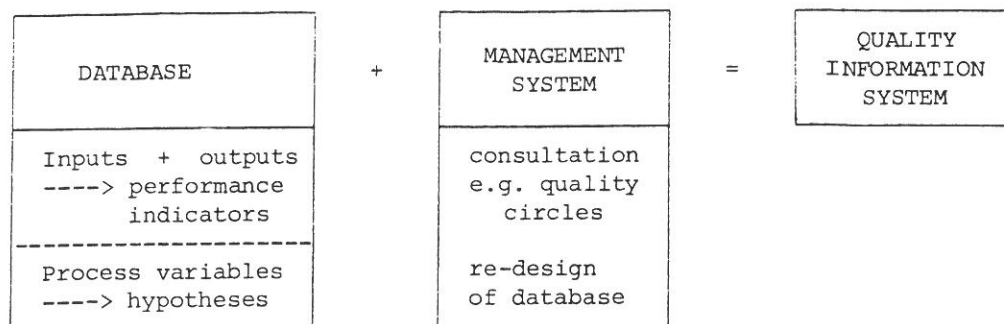
The interpretation which schools might make of this data is not a simple matter. If schools believe, for example, that the examination results of TVEI pupils were damaged by experiential methods, by taking pupils out of classrooms for work experience and residential courses, or by leaving them often in the care of supply teachers whilst their own teachers were receiving INSET courses, what changes should they make? Ignore the warning signs and hope the system recovers from the initial shock of TVEI? Return to traditional didactic teaching and abandon relevance and experience-based learning? Or should some adjustments be made whilst retaining to some extent the new approaches urged by TVEI? Which strategy will improve next year's examination results? We believe these decisions are best discussed by practitioners and those directly responsible for the schools. Quality circles, as in industry, should "take the data on board" and deliberate on the next year's strategy.

The data should also give rise to discussion nationally among those promoting TVEI. They do not need to know which schools or which authorities. They need to discuss the national picture. This provides a good illustration of the differing needs for information at differing levels of the education system which the pupil-level data provides, by aggregation, for all levels.

4. Conclusion

By way of conclusion we present an extension of an earlier figure:

FIGURE 3



We have constructed a research-and-recommend three step approach with an information system strategy and we have urged that the Information System strategy should be managed in a way which provides data to levels of the system at which there is the possibility of flexible responses to the data and professional interest in developing the quality of education on offer. The system must accommodate different goals and must be cautious in interpretation of cause and effect and, therefore, in assigning responsibility. Schools cannot be held accountable for Performance Indicators they cannot influence.

An effective Information System needs technical expertise AND by the involvement of those who will use the data. We have given two examples of databases which have been collected by approximately one hour's testing time per year in schools (not an undue incursion into instructional time). Sometimes these Information Systems ring warning bells and perhaps, like fire fighters, that is all we can expect.

REFERENCES

- Aitken, M. and Longford, N. (1986) Statistical modelling issues in school effectiveness studies.
Journal of the Royal Statistical Society, Series A, 149(1) pp.1-43.
- Cuttance, P. (1985) Methodological issues in the statistical analysis of data on the effectiveness of schooling.
British Educational Research Journal, Vol. 11, No. 2, pp.163-179.
- Fitz-Gibbon, C., Hazelwood, D., Tymms, P. and McCabe, C. (in press)
 Performance Indicators and the TVEI Pilot
Evaluation and Research in Education.
- Glass, G.V. (1979) Policy for the unpredictable (Uncertainty Research and Policy).
Educational Researcher, 8(9), pp.12-14.

- Glass, G.V. and Smith, M.L. (1979) Meta-analysis of the research on class size and achievement.
Educational Evaluation and Policy Analysis, 1, pp.2-16.
- Goldstein, H. (1988) Multilevel Models in Educational and Social Research. Charles Griffin and Company.
- Goldstein, H. and Cuttance, P. (1988) Guardian, 1st March.
- Harrison, P. (1988) Winning points
Times Educational Supplement 27/5/1988.
- Hedges, L.V. and Olkin, I. (1985) Statistical Methods for Meta-analysis. New York: Academic Press.
- McCabe, C. (ed) (1986) The Organisation of the Early Years of the Technical and Vocational Education Initiative. Multilingual Matters.
- McPherson, A.F. and Willms, J.D. (1987) Equalisation and improvement: some effects of comprehensive reorganisation in Scotland.
Sociology.
- Raudenbusch, S. and Bryk, A.S. (1986) A hierarchical model for studying school effects.
Sociology of Education, Vol. 59, (January), pp.1-17.
- Rosenthal, R. and Jacobsen, L. (1968) Pygmalion in the Classroom: Teacher Expectations and Pupils' Intellectual Development. New York: Holt, Rinehart and Winston Inc.
- Willms, J.D. (1987) Differences between Scottish educational authorities in their examination attainment.
Oxford Review of Education, 13(2), pp.211-232.
- Willms, J.D. and Cuttance, P. (1985) School effects in Scottish secondary schools.
British Journal of Sociology of Education, 6(3).