
EVALUATION OF SCHOOL PERFORMANCE IN PUBLIC EXAMINATIONS

Evaluation of School Performance in Public Examinations

A report for the Scottish Office Education Department

C.T.Fitz-Gibbon
Director
Curriculum, Evaluation and Management Centre
School of Education
University of Newcastle upon Tyne
England
NE1 7RU

'Phone 091 222 6380

CEM Library

95

1

Introduction

This report is a list of Recommendations followed by an Executive Summary. The main report is bound separately. An instructional booklet entitled 'Value-added approaches to Performance Indicators' has also been produced as part of this project. Outcomes from Scottish education have for many years been measured by blind-assessed, curriculum-based, public examinations. The brief for this project was to consider ways in which the resulting data on performance in public examinations could be used for **PERFORMANCE INDICATORS**. A performance indicator can be defined as an item of information collected on repeated occasions to check on the performance of a system. Performance Indicators are becoming a recognised feature of complex systems such as the education service. In a complex system which relies on professional commitments Performance Indicators should be recognised as agreed ways of measuring the extent to which agreed goals are being achieved. The author wishes to express appreciation to the members of the working party for many challenging and fruitful discussions: Jonathon Davidson, Alison Kennedy, Stephen Sharp, Peter Tymms and, particularly, to the two outstanding chairs: Harvey Stalker and Colin MacLean.

Recommendations

[References to parts of the main report are in square brackets]

The first four recommendations concern immediate practice, commending a system which is close to being realized and suggesting some additional areas in which progress can be made.

Continuation of the headstart achieved by Scotland

The creation of the Standard Tables is impressive and their development should continue. Notable features of the tables are:

- (a) the use of examination data, both raw and adjusted for intakes;
- (b) the disaggregation by examination subject rather than the use of only whole-school indicators
- (c) the presentation in three-year sets to allow trends to show
- (d) the making of analyses available for every school in Scotland.
- (e) the linking of an indicator system to the work of the Scottish Examinations Board (SEB). By this strategy the SOED has already created an impressively cost-efficient and timely system.
[Chapters III, VI, p 101]

Addition of value-added indicators for Highers

“Value-added” performance indicators should be incorporated into the Standard Tables for Highers as soon as possible, based on the pilot work being conducted at the SEB in connection with this project.

[Chapter III]

***Ethos of a
PI system***

The most important aspect of a Performance Indicator System is its behavioural impact: does it lead to behaviours which provide a better education service? It is difficult to specify how the most effective ethos can be achieved but the following points may help:

- (a) Viewing Performance Indicators as agreed ways of measuring the extent to which agreed goals are being achieved directs attention to the need to develop agreement as the system develops. There is every indication that the SOED is vigorously attempting to develop the necessary agreement and understanding among EAs and schools.

[Chapter I, II]

- (b) The atmosphere in which the PIs are used should be one of information and collaborative investigation, not surveillance and judgment.

[Chapter II]

- (c) Most emphasis should be placed on providing indicators which are disaggregated to the units of management within schools: generally school departments or their equivalents.

[Chapter II, VI, p101]

- (d) Quality assurance in every department should be the principal aim of performance monitoring, not competition between institutions.

[Chapter I, V, p.76]

- (e) School departments/faculties should receive and understand all the PI data. A transparent and accessible system which is perceived as fair is essential if the information is to be used for improvement. Systems which are not accepted can be subverted.

[Chapter II]

**Training and
Resources**

(a) ***In schools***

A substantial programme of training has already taken place for Relative Ratings and National Comparison Factors. A similar effort will be needed for the value-added approach. (The booklet "Value-added Approaches to Performance Indicators", produced as part of this project, is an attempt to assist in this effort).

[Chapter II, VI]

(b) ***For Inspectors***

HMII and EA Inspectors will need to be able to interpret the indicators correctly and with confidence. This may require the provision of on-going training, especially as indicator systems evolve and become less accessible and more complex. Clear and accurate materials have been produced for use with the Standard Tables and further work will be needed.

[Chapter II, VI]

(c) ***Computerization***

Computer-based data collection and feedback procedures will need continued development and, therefore, resources. Computerized data collection will become more feasible and cost-effective as efforts to install standard software in schools begin to be successful. Work by W.Coyle in Strathclyde has set a standard for clear feedback on computer screens.

[Chapter IV, VI, p99]

(d) ***An interrogatable database***

If schools or Inspectors are concerned about particular indicators they should be able to request a more complex analysis, using multi-level modelling and incorporating more factors. In such a situation schools should also be provided with the residuals for each student.

[Chapter III, IV]

The next two recommendations are made to improve the foundations of the quality assurance measures available for the education system: the external assessments of learning outcomes. The first concerns a way in which skills could be credibly and fairly certified, i.e. a way to give the same attention and professionalism to the assessment of skills as is given to academic work. The second addresses ways to improve the working of the academic examination system.

***Skills
Certification
Centres [SCCs]***

Discussions should be initiated by SOED regarding the possible role of Skills Certification Centres. These would be established in major population centres and specially equipped and staffed to provide tests of a wide range of practical skills: scientific, secretarial, technical, linguistic, construction-related etc. These centres could deal with schools and colleges but also directly with students who have acquired skills on their own or, for example, elsewhere in Europe. Such centres could provide credible and reliable certification 'on demand', allowing students to present themselves whenever they and/or their teachers felt they were ready to demonstrate competence. Since certificates from such SCCs could obviate the need for business and industry to undertake some of their own assessments, business and industry might assist in funding a pilot centre. Discussions with representatives from SCOTVEC, SEB, EAs, RSA, the Careers Service, representatives from Business and Industry, and any other group concerned with skills assessment should be the starting point. The present reliance on teachers to take on assessment roles is not satisfactory. Teaching and certification should be kept separate.

[Chapter III, p.31-37]

***The working of
the examination
system***

In view of the increasing importance of examination results due to their use as Performance Indicators, and in view of the possibility of small biases adding up in the aggregation for departmental-level indicators, the relevant bodies should be asked to consider the following improvements in practice:

- (a) Scripts should be marked blind. Names of the pupils and the centres should not be known to the markers. As long as names and centres are on the papers used by markers, the possibilities of gender bias, social class, ethnic or religious bias cannot be ruled out.
[Chapter III, esp.p.38]
- (b) Stratified random sampling should be considered for use in assigning scripts to markers (practice in Scotland is already better than in England and Wales in that the scripts from one centre are not all sent to the same marker.)
[Chapter III]
- (c) Different methods of assessment (such as internal and external) should be separately reported. Different methods of assessment measure different kinds of achievements and aptitudes. The construction of examinations should reflect the literature on multi-trait/ multi-method research.
[Chapter III, p.31-37]
- (d) The SEB should move further towards making all examinations at a given level equally difficult. The extra difficulty of the mathematics and science subjects is probably serving to deter some students from taking up these subjects at Highers. Reducing the difficulty of these subjects just to the point at which they are no more difficult than other Highers could increase the age participation rates in education and increase supply of school leavers who are numerate and scientifically and technically competent.
[Chapter V, p.69]

The next three recommendations look to new lines of development in the use of Performance Indicators.

***The development
of value-added
approaches for
Standard Grade***

At Standard Grade no suitable prior measures yet exist, but it is likely that there will be a demand for a value-added approach at this stage. Consideration might be given to the use of aptitude measures or the use of a General Paper at S-grade. Classroom administered assessments, as currently envisaged, do not promise to be sufficiently reliable and valid for use in an indicator system.

[Chapter III, IV]

***The development
of other
performance
indicator
modules***

Following the anticipated success with the Standard Tables, the question will arise as to what other modules should be developed for the indicator system in education. It is recommended that levels of teacher satisfaction should be considered.

[Chapter VI, p100 ff]

***A parallel
research
programme***

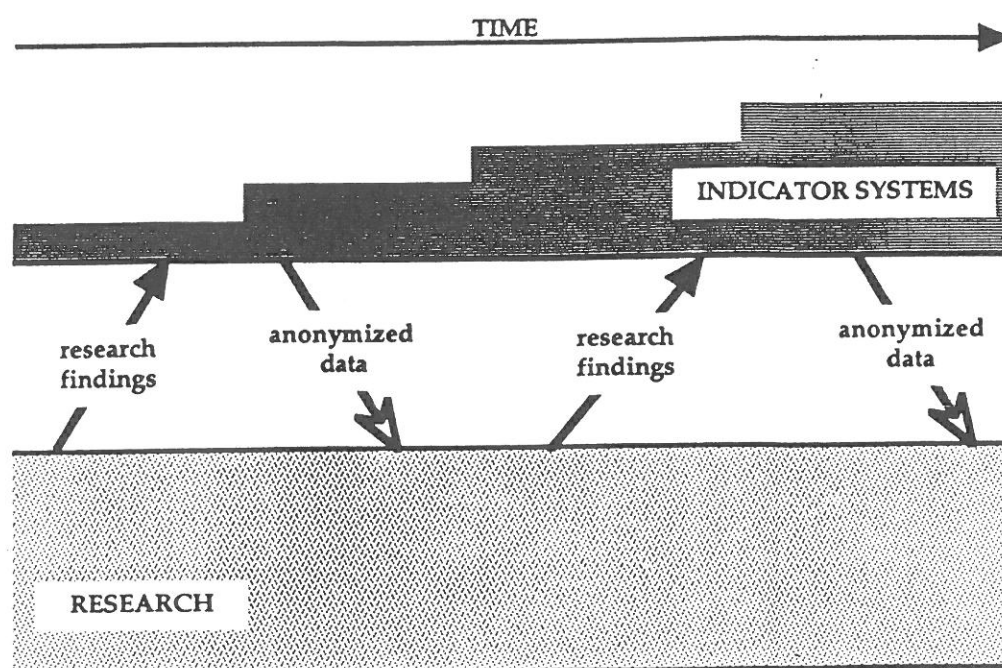
Given the highly valuable data which is collected and analysed in order to provide yearly feedback to schools, it would be wasteful not to maintain and build a research capacity to work in tandem with the performance indicator system. Monitoring systems have been developed on the basis of research findings and research has benefited from the data collected in monitoring. The two activities, research and monitoring, are distinct but each needs the other. (Figure below to emphasise the point.)

Therefore research programmes should be created and maintained in parallel with the monitoring. Some issues which might be urgently addressed are:

- the extent to which multi-level modelling indicates a need to change to multiple indicators for different ability groups or SES groups or ethnic groups (ie the existence and stability of heterogeneous slopes).

[Chapter III, pp25-27]

- how presentation of information about examinations can be matched to user preferences and how the flow of pupils through education can best be tracked and represented.
[Chapter V, VI p99 ff]
- how new modules of the indicator system can be developed
[Chapter VI, p100]
- the impact of having teachers take on assessment roles
[Chapter III, 3.7]
- the impact of indicators on schools, particularly their "behavioural impact".
[Chapters I, II, III]



Summary of the Main Report

Chapter I **Introduction**

Accepting that examinations are an “authentic”, valid measure of an important part of the work of a school, there are two reasons for using examination results for Performance Indicators (PIs):

- (1) to insure that all pupils have *equal opportunities* to achieve well in examinations regardless of the school attended. This requires *quality assurance* in every department of every school. PIs must be used to monitor this quality in a way which is fair to pupils and teachers.
- (2) to provide *feedback* to school departments so that they can monitor their own effectiveness from year to year.

Chapter II **Criteria for** **Indicators**

Performance Monitoring could go wrong. The aim of this Chapter was to create a set of criteria against which proposed indicators could be judged. This evaluation of indicators can help to avoid some pitfalls and to clarify the uses and limitations of each indicator.

The criteria developed arise in part from the literature and recent experiences but can be seen as very largely a matter of logic:

The aim of monitoring is to improve the system. If anyone is to take actions in response to indicators then the indicators need to *refer to the units of management* where such actions might be taken (criterion 1) and the focus of the indicator must be *outcomes over which the units do indeed have some control*. (criterion 2). Indicators which do not take account of the circumstances in which schools work cannot be interpreted and are therefore of limited informational value. They are also potentially unfair. Thus indicators need to be *contextualized* (criterion 3). Information cannot be acted upon if it is not received, therefore all indicators should be *fed back* to the responsible units of

management (criterion 4)*. Information might be rejected or ignored if it is perceived as unfair, so *fairness* is a major consideration (criterion 5) For statistical indicators the fairness will be apparent if the indicators are *accessible, explained and impervious to manipulation* (criteria 6,7,8). Being able to *check* on the accuracy of the data is important for many reasons, fairness being one (criterion 9). Initial acceptance of indicators is not enough. It will be important that *indicators are found to respond to improvement* (criterion 10). Finally, *the kinds of actions taken in response to indicators must be educationally sound* (criterion 11). This is obvious but the application of this criterion to the development of indicators has often been neglected. Hence we see reliance placed on an indicator such as percentage pass rate which could have the effect of encouraging teachers to push weaker candidates out of the examinations, to teach to the borderline group, and to ignore the more able. Finally the indicator system *must not consume too much in the way of resources* which have to be taken out of the delivery side of the education service and put onto the monitoring side.(criterion 12).

In summary, Performance Indicators and the indicator systems in which they are used, must be *relevant, informative, fair, beneficial and cost-efficient*.

Chapter III **Applying the** **Criteria**

In this chapter the criteria developed in chapter II are applied to the following indicators:

- Value added residuals based on simple regression
- Relative Ratings as provided in the Standard Tables being produced by the SOED and the SEB
- National Comparison Factors, also found in the Standard Tables
- Residuals based on multi-level modelling
- Whole school indicators
- Examination totals or means (as opposed to subject by subject indicators.)
- Percentage pass rates
- Internal, teacher-given assessments

A completed evaluation sheet for each of these types of indicator is included in the main report.

The major points arising from consideration of the various indicators were:

(1) The Standard Tables will be highly valuable. They will provide Relative Ratings and National Comparison Factors for examination subjects at Standard Grade and Highers, plus, possibly, a value-added measure for Highers. These indicators meet the more important criteria and together will form a basis for internal self-evaluation and for external evaluation and quality assurance.

(2) Different methods of assessment measure different aspects of knowledge, skills and understanding. This is widely accepted and is the current experience of those looking at Standard Grade data and those researching performance nationally. Not only do different measures measure different aspects of human performance, but the method of measurement has a strong impact on the results. This was foreshadowed in the literature on multi-trait multi-method studies in psychometrics. Since different methods measure different competencies and the methods themselves have an impact, it is highly desirable that the results of different methods should be reported separately, as profiles, and that the measures are not allowed to influence each other. These findings also imply that highly standardised procedures are needed if results are to be fairly compared, a point which calls into question the wide range of options provided for testing in classrooms. Separate reporting would enable much needed research on the predictive validity of newly developed assessment procedures to be conducted and would also ensure that some of the assessment was free from the bias which inheres in face-to-face, internal assessment.

(3) The work of the examination boards in the UK has served as a major method of external validation for public education. The examinations undergo changes but the basic methods provide credible, reliable and valid assessments. Unfortunately there remains the problem of the presence of candidates' names on the papers which are being marked.

Names of pupils and their schools often contain information on the

gender, socio-economic status, ethnicity and religious affiliation of the pupils. The extent to which markers are biased by the information is not known but this would seem to be an area where quality assurance has to be embedded in defensible procedures: names should be removed from scripts prior to their being sent out to readers.

Examination Boards were invited to comment on the issue of the removal of names. The main pleas referred to the expense which might be involved and the danger of misattributing scripts. The use of a front page set up in advance of the examination, designed for OMR (Optical Mark Readers) and incorporating a tear-off slip could surely be investigated. Such a procedure might actually increase accuracy and speed up the grade production process.

(4) Skills Certification Centres. If vocational skills, needed by all students not only those on SCOTVEC courses, are to be treated as seriously as academic skills they need to be externally assessed by highly credible and defensible procedures. The creation of Centres which are professionally equipped and staffed to assess practical skills would meet many needs: for credible assessment, for flexible assessment to match flexible learning, for crediting prior learning of the skill variety, for validating skills acquired in non-standard ways and abroad, for re-assessing skills after periods of non-employment. Importantly they would remove the certification burden from teachers, restoring the role of coach as opposed to umpire.

Chapter IV
Input indicators:
what predicts
achievement?

If indicators are to be fair they have to be 'contextualized', that is to say due consideration has to be given to taking account of relevant factors. The statistical implication is that account must be taken of 'covariates'. Possible covariates are considered in this chapter. Prior achievement is the strongest single predictor of subsequent achievement. If, then, Performance Indicators are to be fair, they must take account of prior achievement. Indicators which do this are often referred to as 'value-added' indicators. To create these, a measure of prior achievement is required. If measures of prior achievement are not available, concurrent achievement measures can provide information on the levels of achievement to be expected in various subjects. Measures of ability are less good predictors but generally

contribute significantly to the prediction of levels of achievement, at least at the secondary stage of schooling. Measures of socio-economic status (SES) present difficult problems, both in measurement terms and policy terms. Along with gender and ethnicity, measures of SES are not good predictors of examination outcomes. It will, however, be important to monitor the effects of all three of these 'demographic variables' in order to guard against bias in the system. Checking the system for such bias, in individual schools or in general patterns, should be the yearly responsibility of a research unit working along with the indicator system. The demographic variables should probably be left out of the indicators which are fed back to the school departments lest the implication given is that less progress from certain sectors of society is acceptable.. To be fair to schools, however, the extent to which differences between demographic groups are proving difficult to remove needs to be kept under review in order that realistic and fair evaluations are made. This essential work should be part of the research programme which parallels performance monitoring.

***Chapter V
Relationships
between
Performance
Indicators***

As Scotland is moving towards the use of the Standard Tables which will contain a number of measures of the quality of examination outcomes, department by department, it was important to look at the relationship between these indicators and at some other possible statistical methods.

Using a sample drawn by the Statistics Division of SOED the Value-added Performance Indicators correlated about 0.70 with the Relative Ratings. These positive correlations indicated that departments which were obtaining good results relative to other departments in the school were also tending to be obtaining good results relative to similar departments in other schools. The strength of the correlations supported other evidence which suggests that the major source of differences in effectiveness lies in departments rather than schools. When value-added measures are considered, there may not be 'good schools' and 'bad schools' but, rather, a variety of departments in most schools.

As emphasized throughout the report, Relative Ratings answer the question:

Taking account of the difficulties of the examinations this year, how have the students in this subject done compared with *how well they did in their other subjects?*

If some departments look very good on Relative Ratings there will have to be some department looking not so good.

On the other hand, the Value-added data for S5 answer the question:

Taking account of their performance in S4, how have the students in this department done compared with how well similar students in other schools did in the same subject.

Chapter VI

Examples of monitoring systems

Monitoring can take place at several different levels of the educational system. The closer the level is to the classroom the more detail is of interest. On the other hand, at the level of the EA or SOED it seems likely that light monitoring is in order: a few major outcomes are kept under statistical review, the 'scanning indicators' spoken of in Strathclyde and the 'warning lights' spoken of in Fife. Both Fife and Strathclyde, however, aim to feed information back to schools at least yearly and therefore aim for 100 percent samples.

At the national level there is the internationally recognised work of the Centre for Educational Sociology, Edinburgh University. In addition to their continuing work with Fife, there is the Scottish Young Peoples Survey, (SYPS) which provides data from nationally representative samples (usually 10 percent) with repeated contacts. This work has established an information system for national planning and contributed much to basic research on school effectiveness.. The School Leavers Survey from the SOED Statistics Branch obtains less detail but aims for 100 percent sample, a census activity. The annual reports of the Scottish Examination Board document the unceasing professional work of the examination system, the backbone of any Performance

Indicator system.

A number of EAs have undertaken considerable and extensive work on performance indicators. Fife's work is notable for the existence of a measure of ability and Strathclyde's by the user-friendly, computer-based feedback. Strathclyde are embedding the indicators work in the context of 'quality'. With regard to Quality Assurance and quantitative Performance Indicators it can be noted that statistical quality control procedures are highly developed in industry. Education in Scotland will soon have some statistical quality control made possible by the Standard Tables.

South of the border, the A-level Information System, developing since 1983 and now growing rapidly outside its home base in the Northeast, provides fair Performance Indicators for A-level provision, using the kind of measure for value-added which may be adopted in Scotland for Highers. The Youth Cohort Study is based on a representative sample like the SYPS.

The further developments which are already being promoted include the computerization of some data *collection*, probably feasible when there is some standard software in most schools. (The SCAMP project).

Computerization of data *presentation* is another line of development. Bill Coyle's work is exemplary in the clarity of the graphical feedback. If busy teachers and administrators are to study Performance Indicators for their schools they will welcome the elegant graph, the picture that is worth a thousand words. Unfortunately the information contained in easily accessible graphical presentations can be misunderstood and false impressions gained. Graphs therefore need to meet the criteria set out for Performance Indicators: they must, for example, be contextualized. They should also have implications for actions, i.e. be diagnostic or informative.

Complex systems such as education will need many inter-connected modules of Performance Indicators. One new module which could be considered is the generation of value-added measures for Standard Grade. The range of information included in modules will also be extended and may include ratings made by inspectors (sometimes called 'qualitative performance indicators'.) Parents and pupils are

being directly surveyed in Strathclyde. A Performance Indicator module in urgent need of development is one which takes into account the views and responses of teachers, including levels of teacher satisfaction.

The changing roles of HMI and EA inspectors

The following statement from Strathclyde about the role of their Regional Inspectors is probably similar to the views of the HMII regarding the part of their role relating to evaluation and monitoring:

“The main purpose of the inspectorate will be:

- to undertake surveys of aspects of provision across divisions or across
- the region as a whole.
- to inspect a relatively small number of individual establishments each year
- to identify excellent practice for dissemination
- to identify areas requiring improvement.

The existence of Performance Indicators will have a substantial influence on how these tasks can be accomplished. If ‘excellent practice’ is meant to imply ‘effective’ practice then such practice should show up in the indicators. It is not a trivial problem to identify practices which work in one place (or appear to) and can be transplanted to other teachers in other situations. There will be disappointments; changes in practice may not yield the effects sought on the indicators.

Yet the existence of an indicator system represents a tremendous opportunity, a chance not previously available to guide the delivery of education with the benefit of regular feedback into the system on its effectiveness.