

is likely to be continuing keen competition for resources with other public services and with private consumption. The education service's case for resources is strengthened if it can show that everything possible is being done to make the best use of the resources already provided. The issue for schools and education services is whether they take a defensive line and try to resist the development of indicators or test them out in a positive way and seek to develop them to the advantage of their pupils.

References

- Blank, Rolf K. (1986). *Science and Mathematics Indicators: a Conceptual Framework for a State-based Network*. Council of Chief State School Officers, State Education Assessment Center, Washington DC.
- CIPFA (1988a). *Performance Indicators in Schools: a Contribution to the Debate*. London, Chartered Institute of Public Finance and Accountancy.
- CIPFA (1988b). *Local Management in Schools: a Practical Guide*. London, Chartered Institute of Public Finance and Accountancy.
- DES (1988). *Education Reform Act: Local Management of Schools*. Circular 7/88.
- Peters, T. and Waterman, R. (1982). *In Search of Excellence*. New York, Harper & Row.

From: Levacic, R. (Ed). *Financial Management in Education*
1989 Open University

Using performance indicators: educational considerations

Carol T. Fitz-Gibbon

DO NOT REMOVE

The need for useful information systems

The major purpose of this chapter is to draw out and explain the educational uses of performance indicators: how performance indicators can contribute to improving several aspects of the delivery of education when they are developed with educational considerations to the fore and with due regard to research findings in social science.

The role of performance indicators as a management tool for improvement is distinct from their role vis-à-vis accountability. Educational institutions may be held accountable for the use of the public funds delegated to them under local management of schools (LMS) and part of this accountability may require examination of 'value added' – what has the institution done for the children on whom the money was spent? But if performance indicators were to be produced only as summative gradings of the institution's performance, then a number of unhappy consequences could result.

First, resistance to the system could develop in institutions with poor performance indicators. They might try to fix the performance indicators without actually altering the educational processes which produced them. If the poor performance indicators were combined with no information on how to improve them, then such reactions would be almost inevitable. Cosmetic adjustments to make performance indicators look good would defeat the purpose of the monitoring system and result in time wasted in battles about appearances. It would certainly be vigorously argued that the information cost more than it was worth.

It cannot be expected that the numerical indicators which constitute performance indicators will be universally welcomed. Old conflicts between

qualitative and quantitative approaches to social science will surface with comments like 'Not everything that goes on in schools can be measured on a scale'. Whilst not everything can be measured, this does not mean that nothing can be measured. How measurements can be made which are fair and useful is, then, the topic of this chapter.

We will start with a brief case study from which we will then extract some general principles for the characteristics of acceptable, fair and beneficial performance indicators in education.

A case study: the COMBSE A-level project

The project called Confidential, Measurement-Based Self-Evaluation (COMBSE) started in 1983 with twelve comprehensive schools representing a wide range of social backgrounds. Performance indicators were developed for two of the most heavily subscribed A-levels: English and Mathematics.

Confidentiality was maintained by giving each school a code-name chosen by the school (e.g. COLDITZ, SHUT, ETON) and known only to themselves and the researcher. This confidentiality was probably important in the production of reports and in the schools' decisions to allow the reports to be put into circulation via the School of Education's library at the University of Newcastle.

The term 'performance indicator' was not in widespread use in 1983 and the question which had to be asked was: 'In order to investigate the effectiveness of A-level provision in a school, what "variables" need to be measured?' To-day the question would be posed in terms of 'performance indicators' rather than 'variables' but the term 'variables' is mentioned to draw attention to the fact that what today may be called 'monitoring' is only different in its ongoing, cyclical and collaborative aspects from what was and is *research*. In particular that line of investigation called 'school-effectiveness' research was very close to providing the information needed for the design of information systems (Gray *et al.*, 1983; Reynolds, 1985; Willms and Cuttance, 1985; Fitz-Gibbon, 1985; Willms, 1987).

What indicators would be fair and beneficial in monitoring A-levels? A logical place to start is to consider goals. What are we striving to provide to A-level candidates? Good examination results? Broadening and educational extramural experiences? A congenial environment at school or college? Since all these were considered legitimate aims for sixth-form provision, we tried to collect performance indicators for each of these aspects. Examination results were collected to see if they were good considering the particular student and what should be expected of that student. We asked about participation in extramural activities so that institutions could see how their students' experiences compared with those in other institutions. And we asked students questions to elicit their attitudes to the institution and their

attitudes to the subjects they had been studying. These, then, might be seen as major *output* indicators:

- examination results,
- participation in extramural activities,
- attitudes to school,
- attitudes to subjects studied.

At the time, one indicator in widespread use was the percentage pass rate for examinations. There are many inadequacies in this indicator, its only virtue being, perhaps, that it is simple. But such a virtue is not sufficient. Indicators should also be fair and beneficial. In Figure 19.2 the inadequacies of the percentage pass rate as a performance indicator are summarized.

Having selected some *outputs* which were considered important, the next task in developing a useful set of performance indicators was to collect information on those *inputs* which might be related to the outputs. For example, if we were going to look at examination results as indicators of school effectiveness, we could not fairly compare the results from, say, a group of pupils most of whom had As at O-level with those from a group of pupils most of whom had Cs at O-level. To make fair comparisons you must

Figure 19.1 Example of a less-than-desirable performance indicator

The percentage pass rate is not a fair or beneficial PI for several reasons:

- no account is taken of the kinds of pupils entered. A 70 per cent pass rate might be a greater achievement than a 90 per cent pass rate if a group of not very able students took the examination in the former class. This lack of adjustment for intake is the most serious problem and applies to any raw set of examination results, such as numbers of passes at various levels.
- barely passing is counted in the same way as getting a high grade, with the result that no extra credit is given for higher levels of achievement. The focus is purely on the pass/fail dichotomy. Since we know that the level of a pass is important to employers and selectors for higher education the reliance on a simple dichotomy is unjustifiable.
- the percentage pass rate may push institutions to allow only potentially successful candidates to attempt examinations.
- the percentage pass rate if calculated for an institution ignores the curriculum balance. Not all A-levels are equally difficult. There is evidence to suggest that a D in A-level Mathematics is about as academically difficult to get as a B in A-level English. An institution attracting many candidates into Mathematics might thus be at a disadvantage when percentage pass rates are examined.
- an inevitable question is 'percentage of what?' Vagueness on this allows the indicator to be strategically manipulated. Are adults counted? What of those who take the examination through some other institution? Are re-sits included?

If an indicator like percentage pass rate is used institutions may be tempted to be guided in their advice to students by considerations relating to the PI rather than relating to the students' long-term prospects, educational needs and interests.

either compare like with like or make some adjustments to take differences into account, or, preferably, take both precautions. How inputs can be taken into account for examination results is the topic of Figure 19.2.

In summary, output indicators were assessed because they represented goals of the institution and inputs were assessed to make comparisons of outputs fair and to give a better indicator of 'value added' by the schools.

In order to take account of differences in intakes, that is, to 'control for' differences in intakes, we needed to find characteristics of the students which were correlated with their A-level grades. The best correlate was the average grade a pupil obtained at O-level, hereafter called the *O-level grade point average*. Since the O-level grade point average was a measure of prior achievement, when it was taken into account we were looking at a measure of 'value added': the increase in a pupil's qualifications. However, there were some problems associated with relying on this measure of intake alone.

Probably the most important problem was that O-levels were set to change (to GCSE) so that we would not have comparable data across the years if we relied only on O-level grades as predictors. To obtain comparable

Figure 19.2 Example of a fairer and more beneficial performance indicator

How is a fairer PI for A-level examination results to be calculated? If we knew what grade to expect for a student, then when the A-level grades became available, we could see whether or not the student had lived up to expectations. If many students in one school achieved better than was expected of them, we might consider that to be a positive PI for the school. Even though all students got Es this could be a very good result if one would have expected them all to fail.

How can we know what to expect of a student? Obviously we cannot know accurately for one student but if we consider, say, a few hundred students, we can make some predictions based on the general pattern. To invent a simple example, suppose over several hundred students we found that the pattern relating GCSE grade to A-level grade was this: at A-level a student tends to get one grade lower than at GCSE. We would then have a 'prediction equation':

$$\text{A-level-grade} = \text{GCSE-grade} - 1$$

or

$$A = G - 1$$

The outcome (A-level grade, A) is predicted from the input (GCSE grade, G). The relationship between GCSE grades and A-level grades is not quite so simple but it *can* be represented by an equation like the one above.

Example: in 1988 the best predictor of A-level was the students O-level grade point average (GPA). These O-levels were scores as A = 7, B = 6 etc. The

A-levels were scored A = 5, B = 4 as in the 'UCCA' scale (University Central Council on Admissions). On the basis of 1988 data, the equation which best predicted A-level Mathematics grade from the average O-level grade obtained by the pupil was:

$$\text{A grade} = 1.85(\text{O-level GPA}) - 8.71$$

Thus for pupils who had an average of B at O-level the expected grade at A-level was 1.85 times 6 minus 8.71 which works out at 2.39, representing an average of a D or slightly better.

For English, the regression equation was:

$$\text{A grade} = 1.50(\text{O-level GPA}) - 5.77$$

What grade would therefore have been predicted for A-level English for students with a B (6.0) for an average O-level grade? (Answer at the end of the figure.)

The essence of the idea is that if we know something about a candidate, such as his or her O-level grade, and if we know how O-level grades in general, in all the participating schools and colleges, related to A-level grades, then we have the basis for predicting the A-level grade each candidate might have been expected to get. If the candidate got a grade higher than the one predicted then that is a positive point for the school. If the candidate got lower than predicted, that is a negative point. *The difference between the actual grade a student gets and the grade predicted for the student is called a residual.* Positive residuals indicate better-than-predicted performance. Negative residuals represent worse-than-predicted performance. By averaging the residuals for a school we can see if, on average, a school's results are better or worse than would have been predicted on the basis of the prior achievement of its candidates.

Of course, a particular candidate's over- or under-achievement may be nothing to do with the school but if a pattern builds up of, say, most candidates doing better than expected then it would look like a positive indication. It is certainly a fairer indication than percentage of passes: it uses all the information (not just a pass/fail dichotomy) and enables outcomes to be compared *having taken account of* O-levels. Statistically this process is referred to as 'controlling for' O-levels. In the COMBSE project, the residuals were averaged for each department in each school, thus providing fair performance indicators for the department.

We have used the example of GCSE grades for simplicity, but of course other characteristics of the students could be 'controlled for': home background, ability, prior achievement on all O-level and CSEs rather than on just the subject matching the A-level subject, and so on.

Indeed, several predictors can be 'controlled for' (taken into account) at once. The procedures for doing this are known in statistics as multiple regression analysis.

(Answer to the regression equation question: Students with O-level grade point averages of B = 6.0 would be expected to get 3.23 in A-level in English, on average, i.e. grades averaging to a little better than C.)

data we turned to testing ability. Three different ability tests, each designed for high-ability students such as A-level candidates, were tried: the AI16 from the NFER, Raven's Advanced Progressive Matrices and the International Test of Developed Abilities (ITDA). The last proved to be the most effective predictor.

Another factor which might be thought to influence A-level achievement was home background. We tried out various measures of this based on father's and mother's occupation and educational levels but on the whole no measure of home background was strongly correlated with achievement at A-level. Although the correlation between measures of home background ('socio-economic status', for example) and achievement may be stronger in earlier phases of education, even there such measures are unlikely to be as strongly related to achievement as are measures of prior achievement or ability. The latter measures are therefore the ones which it is most important to take account of as 'inputs'.

Process variables were also assessed in order to see whether they related to outputs. For example, we considered whether or not some styles of teaching were associated with better outcomes and also looked at amounts of homework reported and allocated instructional time. This we view as just a beginning. Seeking adequate measures of process variables, of what schools do, is the most difficult part of the information system but it is the part which might be of most help in the long run. To locate effective process variables (which may be different in each subject area and different with different kinds of students) will require long-term and close collaboration between those who collect and analyse the information and those who use it.

How was all this data actually collected? By personnel from the university going directly into schools and colleges, giving the ability test under standardized conditions and administering a questionnaire immediately afterwards. One visit to each school or college was generally sufficient. We asked for the students taking A-levels in the coming summer to be brought together under 'examination conditions' for 1.5 hours. These conditions were requested partly because of the need to administer the ability test but also because it was better that the questionnaires were answered without discussion among students so that the answers were their own opinions, independent of others. That no school dropped out of the system may well be a reflection of the fact that at least we made few demands on school personnel.

Over the years since 1983, more schools and colleges have joined the project voluntarily and in 1988 a sudden expansion took place as it spread to five LEAs and included nearly 50 schools or colleges. It was also extended to monitor eleven subjects rather than just English and Mathematics.

When LEAs started to support the project financially, the confidentiality aspect changed: some LEA personnel wanted to see the data and know which code-names represented which institutions. (To signal this change the project was renamed in 1988, becoming ALIS: the A-Level Information System.)

Using performance indicators in management

Each year reports were produced which summarized the data on inputs, processes and outputs. The summaries were provided for each department. By perusing these reports heads of departments could locate institutions similar to their own (from the tables of inputs), and then compare their performance on the outcomes (examinations, participation rates and attitudes) with those of the similar institutions.

The tables of greatest interest were probably the tables of residuals for the examination results. For each student a predicted grade had been computed based on information about the student and on the pattern of results in the entire sample. When this predicted grade was compared with the grade the student obtained, it could be seen whether the student had done better or worse than expected. If the student had done better than expected then this yielded a positive residual. A negative residual represented doing worse than expected and a residual of zero represented a performance in line with expectations. By summing up and averaging the residuals in each department it could be seen whether the students in the department tended to have done better or worse than expected. In other words, the 'residuals' (explained in Figure 19.2) provided fair performance indicators for examination grades and they were aggregated at the level of the department, not summed for the whole school. Here it should be noted that a department would get good performance indicators if it had obtained results in line with students' abilities (or better). Since each student counted equally, the implication for schools was in line with desirable professional practice: to care equally for each student. To get good residuals bright students must get good grades and less bright students must get grades appropriate to their ability. Using residuals as performance indicators, rather than using percentage pass rates, meant there was no temptation to implement exclusive entry policies just to boost the indicator. Entry policies would not affect the indicator and could therefore be based on educational considerations rather than being made with a view to the subsequent statistics.

Were the performance indicators in the COMBSE project educationally useful? Because of the 'self-evaluation' nature of the project and because at the time of writing there has been little systematic research on this question, it is difficult to provide an adequate answer. Perhaps some schools or colleges stayed in the project because they found the data useful in ways which have not come to our attention. However, from the uses we have heard of, examples can be given.

1 Some schools frequently look unsuccessful when compared with other schools simply because the other schools have a more privileged intake and consequently get a 'better' set of raw examination results. Parents may then ask questions, wondering if the school is responsible for the poor results. One of the schools in the sample was in this kind of situation. It was one of the few schools offering Further Mathematics and it sent students to

Oxbridge colleges, despite being located in a far from privileged part of town. An interesting observation about this same school was that although it was attracting many more students into mathematics than normal (about four or five times as many as took English, whereas nationally this ratio is about 1.4) and getting very good results considering its intake, it was not recognized locally as particularly good. Lecturers visiting it in connection with teacher training actually gave the mathematics department a 'below average' rating in an off-the-cuff set of ratings collected to look at the reputation of schools and among a dozen teachers, eight could give it no rating at all. The four who could give a rating gave two above-average and two below-average ratings. Gray and Hannon (1986) drew attention to the inability of HMI to notice when a school with a 'poor' intake did well. The existence of fair performance indicators allows such events to be observed and this school in particular found the COMBSE reports useful in its relations with parents and governors and in reassuring and encouraging its own staff.

2 A second example can be given from quite a different kind of school. This school could be described as having been coasting along happily with a privileged intake and many high examination scores. Concern developed when a selection of students' responses to an open-ended question were included verbatim in the annual report. Comments like 'We're just like fifth formers only without the uniform', 'They don't treat us like adults', 'Staff are sarcastic', were worrying, especially when the proportion staying into the sixth was found to be declining. A staff meeting was called to discuss the findings. Following a request for more information from the data base, *all* students' comments were reported back to the school (suitably edited for anonymity) and the tenor was certainly not uniformly negative – which reassured staff somewhat. Nevertheless the performance indicator on the attitude scale was placing the school each year consistently among the two or three schools with the least positive attitudes among its sixth-formers. Plans were made to improve the sixth form with a view to increasing the satisfaction of students.

Now perhaps it doesn't take a data base to tell you if the atmosphere is good in a sixth form, but the data base did add to the evidence and did convince some who thought there was no problem that there might be a problem. Furthermore, the data could be related to ability groups and rule out the idea that it was only low-achieving students who were dissatisfied: dissatisfaction was across-the-board. However, the interpretations made must be cautious. It could be that the kind of students typical of this school, the sons and daughters of professionals, tend to be more critical in the sixth form than the students from more working-class backgrounds. This negative correlation between attitude and parental occupational status did hold in the COMBSE data base accumulated over five years, and this possible relationship is being further investigated as more data becomes available.

3 Another use made of the COMBSE data was to investigate differences in difficulty between A-level subjects. An investigation of the difference in difficulty level of Mathematics and English suggested that a D in Mathematics was indicative of the same level of academic ability as a B in English (Fitz-Gibbon, 1988).

Limitations

Although some instances of use of the data have been provided, the information was less useful to schools than it might have been and the reasons for this need to be considered.

First, the project provided a set of performance indicators, a data base, but did not provide a person to ensure the data were understood or used. One meeting a year was not enough for this purpose, especially when there were no other pressures towards paying attention to indicators.

Another possible hindrance was the nature of the reports: they were heavy and indigestible. There was a good deal of information and much of it appeared complicated at first sight. The mathematics department heads might have taken the report home for light reading but the English department heads, quite understandably, may have largely ignored it. The single best-recalled part of the reports appeared to be the verbatim comments from students.

It might be thought that the tables should be simplified and verbatim comments should be reported back more often, but there are a number of aspects to be considered before such a recommendation could be accepted. Leaving the tables in a somewhat complex form discouraged trivial and precipitous interpretations by uninformed persons. The information system was for the colleges, schools and LEA; it was for a professional rather than a lay audience. Better to concentrate on training managers to cope with the tables than to oversimplify. Furthermore, there would have been costs involved. The tables used were taken directly from the computer output, which ensured there were no transcription errors and was quick, even though some tidying was needed to make them more elegant and intelligible. To re-work the tables further would have increased the costs of the production of the reports.

The costs argument applies acutely to the question of feeding back the verbatim reports. Before they can be fed back they have to be carefully read to remove remarks which betray the institution's or the student's identity. They also have to be typed, which is costly; and they take up many pages, which increases reprographics charges. Nevertheless, such is the interest in these comments that at least a one-in-five sample of the comments should be reported back to institutions. If resources were less limited, formal coding procedures could be used on the entire set of comments. The problem is entirely one of cost.

To evaluate school performance, the indicators must have two

important features: they must be *comparative* and at an *appropriate level of aggregation*.

Much of the usefulness of the data derives from its *comparative* nature. Take, for example, just the responses to the invitation to agree or disagree with the statement 'I would recommend to others that they take their A-levels here'. What percentage agreeing with the statement would be acceptable? Thus even simple figures need to be considered in three ways: (1) Is the figure satisfactory taken at face value? (e.g. is 70 per cent positive response acceptable?) (2) How does the figure compare with that from other schools like ours? (3) Does the figure need to be adjusted for intake differences? The last two questions can only be answered with a comparative data base across institutions.

The *level of aggregation* must be appropriate. Much of the discussion of performance indicators has been couched in terms of 'good schools' and 'poor schools', but is it not the case that the same school may be good or poor in different ways? Take examination effectiveness, for example. Good and poor departments exist in the same institution. Adding up the residuals given to each department, in order to obtain a school average, would simply obscure the information about the more and less effective departments and give no indication of which departments the school might most appropriately seek to improve. Much more research is needed before we know if there is a 'school effect' which influences work in every department, and meanwhile it would seem wise to aggregate only at the level of the department, not up to the school level.

Discussion: lessons from the case study

Case studies are illustrative; they do not establish proof that the lessons drawn from the individual case will generalize. Nevertheless, it is useful to try to see some general patterns in a case study which might apply to other situations. In the following pages we will look at characteristics which, partly on the basis of the experience gained with the A-level Information System, seem to be important. These desirable characteristics will be discussed under headings which reflect the steps you might use in developing an information system (see Figure 19.3).

Approach

If LEAs are to make a success of the monitoring role which the Coopers & Lybrand report and the DES circular 7/88 suggested, then their approach to the institutions they will monitor needs to be carefully considered. The view put forward here is that the approach needs to be very much in the spirit of *collaboration* with schools and colleges: collaboration in information seeking and interpretation, with the overall goal of improving the delivery of

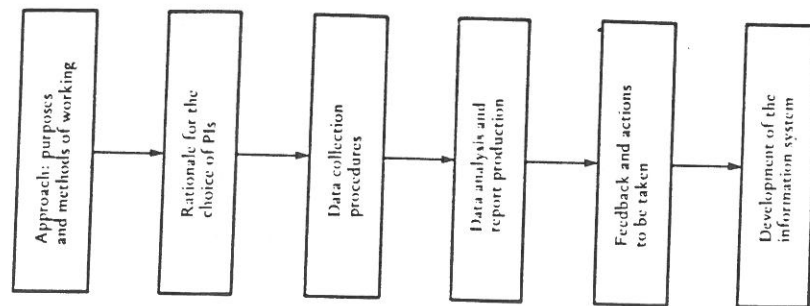


Figure 19.3 Steps in developing an information system

education. This kind of approach is strategically advisable for reasons argued at the beginning of this chapter but it is also, quite arguably, the only intellectually defensible approach. Not a great deal is known about the effects which schools have on pupils, about how much effect schools can have on attitudes, achievement, 'demeanour', delinquency rates and so on. In the absence of a strongly established research base, the monitoring must be investigative, not conclusive. Schools cannot be held responsible for outcomes over which they have little control: no accountability without causality! But little is known about which outcomes can be directly affected by schools, and even less is known about what changes to make in order to improve the outcomes. In the face of this situation no strong-arm judgemental system is defensible.

Rationale

Not every number which spews from a computer or a budget is worthy of the term 'performance indicator'. A performance indicator should have some observable and arguable, if not proven and evident, relationship to performance. Performance indicators are taps into a model of how the system is working. Thus, in the COMBSE project, outcomes were selected which were consistent with widely accepted goals. Then input indicators were located which needed to be taken into consideration because they had a bearing on the output indicators. Process indicators were selected to try to find actions which the schools took which might relate to *how* the schools' goals were attained. The need for there to be comparative data which has been derived from pupil-level data suitably aggregated was stressed above and the criterion added here is that the indicators should be chosen because they are elements in an explanatory model of the system being monitored.

Data collection

The data collection method should be designed to produce credible and defensible data of research quality while making few demands on staff and students. In the COMBSE study, data were collected specifically for the information system, by persons independent of the school. This represented a cost but seemed worthwhile in terms of credibility, convenience and efficiency.

If schools were asked to administer their own questionnaires about students' levels of satisfaction would the data be seen as credible? Would the students believe an offer of anonymity if their own teachers were collecting in the questionnaires? It might be argued that the questionnaires could be completed anonymously but that would mean that the questionnaire data could not be linked with examination data. One could not, for example, investigate whether positive attitudes were expressed by a range of students or only by students who were doing well academically.

Another problem which could arise in the questionnaire administration is the building up of a group ethos by comments spoken aloud when the questions were being answered. There might be some mileage in having students discuss the attitudes in groups but a few vocal students can sway the opinion of the group quite considerably. It seems safer to aim to get each pupil's individually considered response, at least in the first instance.

Then again, the atmosphere set by the way the task is presented and explained to students could influence their responses on the questionnaire. In the COMBSE project the explanations were put on a tape and this was played to each group. Thus all participants heard the same explanation in the same tone of voice.

It is not only attitude measurement which needs careful attention to the conditions of administration. The tape we used also timed the test. It takes

schools no time at all to realize that they can improve their residuals based on controlling for ability by obtaining low scores on the ability test, making it appear they are working with less able pupils than is in fact the case. Simply shortening the administration time could provide a boost to such residuals.

Data analysis

The first step in data analysis is to ensure that the data are accurately entered into the computer and are being read accurately, so-called 'data cleaning'.

As a further check on the accuracy of the data it is important to build in some checks which can be made by schools and colleges. In COMBSE the raw unadjusted mean A-level score was reported. If our figure was different from the institution's figure the problem could be investigated and was usually due to absence at the time of the administration of the test and questionnaire. One point on which there is certainly consensus is that the basic achievement and attitude data should be collected and analysed *pupil by pupil*. School-level aggregates can be seriously misleading as well as uninformative about the details (Aitkin and Longford, 1986; Woodhouse and Goldstein, 1988).

Feedback and actions consequent upon it

Feedback is an important concept both inside and outside social science (as in the entire science of cybernetics). There is some evidence that feedback in and of itself is motivating, without the need for contingencies dependent upon the level of performance.

Quality circles, a concept familiar in management, would seem to be the way to ensure that indicators are examined and, where appropriate, actions are taken. A quality circle might be established within each school with a brief to: evaluate a department's performance taking into account the PIs on outcomes which are valued; consider what actions are needed in view of the evaluation; if improvement is sought, to consider the process variables which have been found to correlate positively with residuals and ask if any of these might be adopted or, if already used, exploited further; to consider explanations for the PIs and where appropriate suggest how information in the data base might be used to check the hypothesized explanation. Initially, someone may need to work with school personnel to develop these activities.

There were two pieces of information on the COMBSE questionnaire which could be directly related to the costs of providing an A-level course: the minutes per week of instruction and the class size. Comparative data on these variables can help managers to see if their courses are very much more costly or very much cheaper than courses provided elsewhere. They will also want to know if the residuals relate to minutes per week or to class size. Adding this kind of information to good educational judgement might assist managers in

reducing costs without damage and/or in putting money where it is needed to improve results.

Development

It has been argued that a data base needs a human interface; it needs to be part of an information system. People are needed who sit down and talk through the data with management teams at each school, who explore possible explanations with the management teams, who listen to ideas from management teams for other data which should be collected to explore other facets of the education system being monitored, facets which might explain the indicators.

The lack of a management system to facilitate the flow of information into the data base (e.g. revisions of the questionnaire) and the flow of information from the data base to schools in a way which made it useful, was a serious deficiency in the COMBSE project and highlights an important point: management information systems require more than performance indicators and information. They require people to make them operate effectively. The most likely people, it seems, are LEA advisers whose new roles include being inspectors as well as advisers.

Research

A monitoring system produces huge amounts of very valuable data. It would be sheer profligacy not to ensure that the data base is mined for valuable information. In addition to the examples which have already been given (the difficulty level of examinations in different subjects, the difficulties of the same subjects with different examination boards, the extent to which schools can actually influence PIs, no accountability without causality, . . .) there are many other questions. There are many opinions about formal, didactic, experiential, pupil-centred approaches but few empirical studies. And what of the effectiveness of schemes like supported self-study or other ventures into distance learning? Are some textbooks ideal for some syllabuses? Or for some pupils? What is the effect of students' having jobs outside school? There are many questions which participants may think of each year. Someone should attempt to answer them in collaboration with the institutions.

Who will do this research? During the heady days of development of the information system you can probably find university or polytechnic researchers only too keen to participate in research and development. Eventually perhaps every LEA will have a thriving R&D unit. Indeed, after the LEA has distributed the money according to the formula, then goal setting, monitoring, evaluating, and research and development may be *the* role which is left. It is a very important and beneficial role if undertaken in a spirit of collaborative research.

References

- Aitkin, M. and Longford, N. (1986). 'Statistical modelling issues in school effectiveness studies', *Journal of the Royal Statistical Society. Series A*, 149(1), 1-43.
- Coopers & Lybrand (1988). *Local Management of Schools*. London, HMSO.
- Fitz-Gibbon, C. T. (1985). 'A-level results in comprehensive schools: the Combse project, year 1', *Oxford Review of Education*, 11(1), 43-58.
- Fitz-Gibbon, C. T. (1988). 'Recalculating the standard', *The Times Educational Supplement*, 26-8-88, p. 15.
- Gray, J. and Hannon, V. (1986). 'HMI interpretation of schools' examination results', *Journal of Educational Policy*, 1(1), 23-33.
- Gray, J., McPherson, A. F. and Raffie, D. (1983). *Reconstructions of Secondary Education Theory, Myth and Practice since the War*. London, Routledge & Kegan Paul.
- Reynolds, D. (ed.) (1985). *Studying School Effectiveness*, London and Philadelphia, The Falmer Press.
- Willms, J. D. (1987). 'Differences between Scottish educational authorities in their examinations attainment', *Oxford Review of Education*, 13(2), 211-32.
- Willms, J. D. and Cuttance, P. (1985). 'School effects in Scottish secondary schools', *British Journal of Sociology of Education*, 6(3).
- Woodhouse, G. and Goldstein, H. (1988). 'Educational performance indicators and LEA league tables', *Oxford Review of Education*, 14(3).