MORE ON EDUCATIONAL LEAGUE TABLES.
HOW DIFFERENTIALLY EFFECTIVE ARE DIFFERENT TYPES OF 'A' LEVEL PROVISION?

Antony Fielding
University of Birmingham

Much has been written in criticism of school league tables in recent years. They have been the focus of a Radical Statistics broadsheet (Radical Statistics (1992)). Yet they are still with us. The latest in the round have been attempts to compare the 'A' level performance of Sixth Forms in schools with that of Sixth Form Colleges, Further Education Colleges or Tertiary colleges. The performance indicator in question was the usual mean 'A' level points score for an institution across all individuals entered for two or more 'A' levels. Reporting on their publication in November 1993, the Guardian (November 24, 1993) was led to conclude, _The overall performance of more than 430 further education and sixth form colleges based on raw exam results is weaker than that recorded for all schools .... Only two sixth form colleges better the average of 20 A-level points for candidates taking two or more A-levels. That is some way behind the leading state selectives who scored up to 30 A-level points'. State selectives apart, then, it seems there is a continuing impression that to do well, where possible students should stay at school and avoid transferring to the more adult atmosphere of post-16 colleges. Ultimately I will present the basis for some league tables of my own which may reverse this impression.

There have been two main general foci to overall critiques of league tables. In the first place there has been little attempt until recently to take into account input and other process factors in producing the raw mean performance indicators. The point has been made that concern should be with value-added. Again the aggregation process ignores individual students and it is surely important to highlight the effectiveness of institutions as they pertain to particular students with unique needs and characteristics. Concentration on the single institutional level obscures those individual comparisons in the overall averaging. Amongst others, Gray and Hannon (1986), Nuttall et al, (1989) and Willms (1992) have drawn attention to inappropriate conclusions that can be drawn if input and process are ignored. Again, for example, Aitken and Longford (1986), Bryk and Raudenbush (1987), and Goldstein (1987) have drawn attention to the consequences of ignoring the hierarchical nature of the educational process in attempting to assess effectiveness. This has led to suggestions that multi-level models be used in data analysis and the methodological and educational literature now shows that such approaches are to be preferred. The Radical Statistics broadsheet referred to previously summarises all these pitfalls of the official league tables admirably.

Of course individual level data of the sort required for a multilevel approach are not always available. That they should be is self evident and points to the need for such data to be focused on as a matter of routine if effective performance indicators or to be developed. However, in the absence of such data Thomas, Nuttall, and Goldstein in a series of Guardian articles (e.g. October 20th 1992) have demonstrated an approach which seems to some extent to control for input. Briefly, institutions are banded into groups on the basis of average GCSE scores for pupils taking A/AS levels. Then for each group a 'regression adjustment' is made to allow for the fact the average A/AS level score increases with GCSE score. The institutions are then placed into bands according to whether A/AS score is better, similar or less than that of institutions with comparable intakes. Quite different judgements of the effectiveness of institutions then ensue. However, the writers are careful to point out that their 'value added' tables are not motivated by any desire to create alternative league tables. They point out the intention to illustrate that DFE league tables will be misleading concerning the real effectiveness of schools and colleges, so that, therefore, they are best ignored.

There is some indication that the DFE is taking note of some of these constructive suggestions and are instigating research into how proper attention to value-added criteria may be made. There is, however, no indication that they are extending this to allow for crucial social-economic factors or other educational process characteristics. Progress between one set of tests and another will also heavily depend on some of these factors. Some early indication of these new DFE approaches were reported in the Daily Telegraph (August 10, 1994) who commented that it was the first time they had 'poked their toes into these waters'. The tables reported comparisons of type of institution (independent, grammar (GM), grammar (LEA), comprehensive, sixth form college, further education college) in terms of percentage of candidates obtaining 20+ points at 'A' level. The comparisons were broken down into 9 groupings by GCSE points score for individuals. No comparison of individual institutions have been made. A fuss was made in this published report about the obvious higher performance of independent schools in all GCSE bands. Other process factors have of course been ignored. There was nothing much in the way of striking differences on the basis chosen between the other types of institution in all the bands, so this was not really commented upon. Yet five days earlier I find a report in the Times Higher Education Supplement (August 5, 1994) reiterating DFE claims that 'pupils in schools get the best results'. The basis here was percentage of candidates getting 25 or more 'A' level points. There was nothing in this report to indicate that the DFE was giving any cognisance to its banding exercise. Is this really further evidence of selectivity of statistical evidence to suit whatever point the DFE wishes to make to a range of requirements from far from disinterested reporters?

Apart from the general critiques of official league tables outlined above there is a more specific point which commentors from the further education sector frequently make. This focuses on the crucial differences in opportunities offered by the different types of institution. In a letter in response to the THES report a director of a technical college (THES, August 12, 1994) comments on the range of choice of qualifications being greater in FE colleges, less good in sixth form colleges, and worst in schools sixth forms. Success rates in BTEC courses are higher. Other commentators have noticed that students often offer a mix of post-16 qualifications and comparisons based on A-level points scores are thus

misleading. The further point made is that various league tables are based on percentages of entries, not initial course enrolments. Since there is a selection effect and also higher drop out rates of pupils in schools (or sixth form colleges) the 'A' level results as currently presented will be biased in the latter's favour. In reporting my own work to follow I meet these comments to some extent by focusing on effectiveness in single 'A' levels and also include all students enrolled for at least one term on a course in comparisons made.

Another concern about the official choice of performance indicators is their lack of any linking to resource factors. This is despite the fact that there has been a growing governmental concern with cost-effectiveness as a way of appraising public policy in many areas, and in particular education including 'A' level provision. 'Cost effectiveness', 'efficiency', 'value for money' are phrases which are pervasive in official comments or published studies. The series of Audit Commission reports in the eighties on the uses of resources in schools or colleges frequently focuses on these issues as do white papers such as Better Schools (DES (1985)). Uses of resources through local management of schools permeate the 1988 Education Reform Act. A link is constantly being made between costs, management, and effective schools. Policy makers are requiring assurances that resources are being used efficiently as well as effectively, so that cost-effectiveness of educational provision now has to be demonstrated. The transparency of all this through league tables based on performance indicators is hidden. As yet, despite governmental self professed concern, there have been few attempts to undertake cost-effectiveness studies of the sort that might inform the concerns that have been expressed. Links of performance to costs and other factors have not, at least publicly, yet been made. Perhaps the real reason is the lack of an adequate database to implement such concerns or possible where data is available an unwillingness to implement the ideas which might lead to conclusions contrary to conventional wisdom.

My present exercise reports briefly some results of a reanalysis of some data originally collected in a cost-effectiveness study by Thomas (1990). The data set is a particularly rich one collected in the early 1980's but I believe has some contemporary relevance in pointing to the sort of routine data that should be collected and used and to some of its conclusions. It also offers scope for application of multilevel methods. The source was twelve institutions in an area of England, six sixth forms in 11/18 mixed comprehensives, three sixth form colleges, two further education colleges, and a tertiary college. Three consecutive cohorts were investigated and information was collected on students, teachers, and institutional characteristics of all subject teaching groups in these cohorts. Full details of all the relevant variables, their nature and sources are given in Thomas (1990) or Appendix A of Fielding (1993): A copy of the latter is available on request.

Thomas recognises three perspectives to cost-effectiveness according to who was bearing the cost; the individual student, the providing organisation (LEA), or society. Some fairly complex calculations went into identifying these various costs, but at the end of the day Thomas derives for each institution a cost effectiveness ratio. His unit of analysis was the teaching group so that per capita costs and per capita outcome score in terms of mean 'A' level score relate to this unit. Weighted averages for these per capita figures produced similar figures for each institution and from these the ratio was found. Attempts were made to control though an analysis of covariance for 'O' level qualifications of the candidates in a teaching group and other process factors, number of teachers

who taught the group, size in terms of number of _A' level candidates, class hours devoted to the course, and dummies to represent eight different subject groups. Full details of the ramifications and detailed conclusions are available to the interested reader in the original report. However, of interest here are institutional rankings outlined in Table 1.

Table (1) Ranking of cost-effectiveness ratios by institution

(Source: Thomas (1990))

	Earnings foregone criterion		Institutional costs criterion		Social costs criterion	
	By final standards	By adjusted outcome	By final standards	By adjusted outcome	By final standards	By adjusted outcome
Institution SFC1	1	1	1	1	1	1
SFC2	8	3	5	3	4	3
SFC3	2	2	6	5	3	2
SF1	11	11	11	11	11	11
SF2	5	5	4	7	2	4
SF3	3	3	9	10	5	8
SF4	12	12	12	12	12	12
SF5	10	10	10	9	9	9
SF6	9	9	7	8	7	7
FE1	7	7	3	2	8	5
FE2	6	6	8	6	10	10
TC	4	4	2	4	6	6

In this table the 'by final standards columns' are cost-effective ratio rankings before any controlling operation. The 'by adjusted outcome' columns represent rankings of actions after the analysis of covariance has produced adjusted outcomes. The labelling of the institutions should be self evident. The move from final standards to a 'by adjusted outcome criterion, in each case produces a fairly clear pattern. Even before adjustment in each case there appears to be no clear advantage attached to school sixth forms. However, after adjustment, and with with one exception the final column, it appears that school sixth forms are the least cost effective. As Thomas states, 'Yet the general impression remains that there is a pointer to some kind of institutional break at 16+ as a contributory factor to the cost-effectiveness in the case of 'A' level provision.' Thus these results stand in stark contrast to the rankings and impression that might have been produced from the raw mean score league table indicators, even if some elements of banding had been introduced.

The analyses just discussed have used to some extent an aggregate approach based on the teaching group as a unit of analysis. They have met the desire to incorporate cost effectiveness into value judgements and have also incorporated an element of the value added approach. However, as previously suggested,

there seems to be a compelling case for a hierarchical analysis. For these reasons a data reorganisation was undertaken and a multilevel approach to the analysis conducted. The restructuring of the data from the three distinct teaching group data sets proved a time-consuming and by no means easy task but eventually it was possible to produce a data base amenable to analysis by ML3E (Prosser et al (1991)).

Again the three perspectives are examined. In the individual perspective ACRHEF is the ratio of individuals 'A' level score on the one subject (zero for drop-outs and failures) to his/her earnings foregone. Timetable variables entered into all these calculations in a fairly complex way. This could then be used as the response variable in the analysis. From the institutional perspective costs can be allocated to a teaching group under three main headings, institutional overheads, capitation and salary costs. Capitation data was missing for a large number of teaching groups but capitation was in fact a very small proportion of overall institutional costs so it was decided to ignore it in forming the criterion. Salary was missing for some groups so that after listwise deletion, analysis was conducted for 9940 individual course members out of an original 10685. A figure for per capita institutional costs can be calculated in ways indicated in Fielding (1993). The resulting effect to cost ratio was formed as INSCR for each individual. Again after performing a necessary calculation to apply social cost to each individual a ratio SINSCR may be formed. A forthcoming paper will present the full details of the variable formation, estimation of detailed effects, other aspects of the analysis, and a more complete account of conclusions (Fielding (1995). For the present purposes I present a few examples of institutional rankings arising out of the analyses in Table 2. In the main since there were only 12 institutions these were treated as a fixed effect dummies. The analysis was thus two level with individual students at level 1 and teaching group at level 2. Also intercept was treated as random at the two levels although in some further work some explanatory variables such as 'O' level score of student were allowed to be random at level 2. The effects on institutional rankings were negligible although the results do suggest some interesting explanations. Although sample size cautions against it we have also tried some three level analyses with additional level three explanatory variables. Again broadly similar conclusions about the placing of institutions were noted and detailed results are not discussed here.

In each of Tables 2(a)-2(c) referring to the three cost perspectives column (1) is the unadjusted two level analysis with no explanatory variables apart from the dummies for institutions, column (2) introduces individual's 'O' level score, column (3) adds to this some teaching group variables, average 'O' level score of group, number of teachers who taught the group, number of 'A' level candidates, and a series of dummies to account for eight subject groups. These are similar to the variables applied by Thomas in his analysis of covariance at the teaching group level. The stimulus behind rows (2) and rows (3) is the differentiation between Type A effects and Type B effects discussed by Willms (1992). He claims that Type A effects should include all effects of institutional policy or practice, composition (e.g. aggregate characteristics of students), composition (e.g. size), or other local socio-economic factors. Thus we control for student input. In contrast Type B effects include only policy and practice variables and control for other relevant ones. It might be noted that where group mean 'O' level score was included it had a negative coefficient, a phenomenon noted in other effectiveness studies.

It will be noted that once we begin to control for input and then some teaching group characteristics the rankings begin to show some stability and are broadly are consistent with the earlier Thomas results. The tables really require no further commentary but do show a contrast to conclusions about placing of school sixth forms arising out of the current league table exercises. It is true that this data is somewhat dated but it would surprise me if a great deal has changed under the current new arrangements for financing the post 16 sector. I do believe, however, that the results and analyses are a basic pointer to the full requirements of comparative exercises and also to modes of analysis. As the radical statistics broadsheet says 'league tables of raw scores are misinformation. They show nothing about the quality of teaching for each school, nor about the breadth of curriculum offered, nor about the suitability of the education offered for the pupils who enter.' I would add that they show nothing about resource allocation and its management. It is clear that different value judgements about what is important to publish will lead to different and loaded conclusions about the effectiveness of different types of institution.

Table 2 Institution ratings under various analyses

	(a) Earnings foregone perspective			(b) Institution cost perspective			(c) Social cost perspective		
	(1)	(2)	(3)	(1)	(2)	(3)	(1)	(2)	(3)
SF C1	1	1	1	1	1	1	1	1	1
SF C2	4	3	4	5	5	3	6	5	4
SF C3	2	2	2	3	4	2	3	2	2
SF 1	12	9	8	11	8	10	12	8	11
SF 2	5	8	9	4	9	9	4	9	7
SF 3	8	12	7	10	12	12	9	12	12
SF 4	7	11	12	8	10	8	7	11	8
SF 5	9	10	10	9	11	11	11	10	10
SF 6	11	7	11	7	6	7	10	7	9
FE 1	10	5	6	6	2	6	8	4	5
FE 2	3	4	3	2	3	4	2	3	3
TC	6	6	5	3	7	5	5	6	6

References

Aitken, M. and Longford, (1986). Statistical Modelling in School Effectiveness Studies, Journal of the Royal Statistical Society, A, 149, 1-43.

Bryk, A.S. and Raudenbusch, S.W., (1987) Application of Hierarchical Linear Models to Assessing Change. Psychological Bulletin, 101 (1), 147-158.

Department of Education and Science, (1985). Better Schools, CMND, 9469, London, HMSO.

Fielding, A., (1993). Institutional Disparities in the Cost Effectiveness of GCE 'A' Level Provision: A Multilevel Approach to the Student Perspective. University of Birmingham Department of Economics Discussion Paper, 93-10.

Fielding, A. (1995). GCE Advanced Level; Institutional Cost Effectiveness, Economics Education (forthcoming).

Goldstein, H., (1987). Multilevel Models in Educational and Social Research, London, Griffin.

Gray, J. and Hannon, V. (1986). H M I's interpretation of School's Examination Results, Journal of Educational Policy, 1(1), 23-33.

Nuttall, D.L., Goldstein, H., Prosser, R., and Rasbash, J. (1989), Differential School Effectiveness, International Journal of Educational Research, 13, 769-76.

Prosser, R., Rasbash, J., Goldstein, H. (1991), ML3, Software for Three-Level Analysis, User's Guide for V.2., London, Institute of Education, University of London.

Thomas, H., (1990). Education Costs and Performance, A Cost-Effectiveness Analysis, London, Cassell.

Willms, J.D., (1992). Monitoring School Performance: A Guide for Educators, London, Falmer.