

The Sixth *QHE* Seminar

The End of Quality?

Birmingham, 25–26 May, 2001

Teaching quality assessment scores: measuring quality or confirming hierarchy?

Dr Lynn Drennan and Professor Matthias Beck

Glasgow Caledonian University

Abstract

Utilising data from the *Times Higher Education Supplement* league tables for the United Kingdom universities, covering the years 1993 to 2000, this paper examines the relationship between teaching quality assessment (TQA) scores, reputational factors and resourcing indicators. The analysis identifies a significant correlation between TQA scores, student entry standards and research assessment exercise (RAE) results. This is further confirmed by a principal component analysis which reveals a high multi-collinearity between these variables, such that the reputational factors (student entry standards and RAE results) exhibit the highest loadings. The conclusion is that, rather than representing an independent indicator of teaching quality, TQA scores largely serve to confirm existing hierarchies, which may be driven by factors other than teaching.

Introduction

The introduction of the Further and Higher Education Act in 1992, led to the formation of new universities, referred to in this study as ‘post-1992’ institutions. Alongside this expansion of the university sector, came a consequent desire to ensure that standards of teaching and learning did not fall. Quality control, audit, assessment and enhancement were now a regular feature of academic debate and discussion (Morris, 1995; Frazer, 1992). Supporters of this new quality movement welcomed the opportunity to re-assert teaching as a valuable activity within a university. However, early on, critics pointed to weaknesses and biases within the new assessment system.

Barnett (1994), who had analysed the outcome of the first funding council quality probes, in England and Wales, for instance, found that the list of institutions achieving ‘excellent’ scores in the teaching quality assessments were highly correlated with the élite institutions in the UK, that is, the traditional ‘old’ universities. This led him to suggest that a cross-subject set of criteria was operating, which favoured certain kinds of institution, despite the Higher Education Funding Council for England’s (HEFCE) intention to be mission sensitive.

Teaching quality assessment

While the exact style of TQA varied between the four parts of the UK, they all encompassed common elements. This included institutional or departmental self-assessment, the production of a self-assessment document and a visit by a team of academic peers. Most of the assessors were nominated by the higher education institutions themselves, with the inclusion of some industrialists or practitioners.

From the first full round of the TQAs in 1993–94, it became apparent that the older established universities were gaining the largest percentage of the higher grades of assessment, with the new institutions trailing behind the rest. This development was attributed to a number of factors, ranging from preconceptions among assessors recruited from old universities to the under-resourcing of many new universities.

Data sources

Teaching quality assessments have taken place annually since academic year 1993–94. The data used for this study is that produced by the funding councils and published in the *Times Higher Education Supplement* league tables for the year 2000 (THES, 2000). As each of the universities is multi-disciplinary, they were subject to a number of TQAs across a wide range of cognate areas. Unlike the TQAs, which are organised by each of three regional funding councils in England, Scotland and Wales, and the Department of Education in Northern Ireland, the research assessment exercise (RAE) is a UK-wide undertaking. The first RAE took place in 1992, when the new universities had only just come into existence. There was, therefore, limited involvement in the 1992 RAE by these institutions and it was not until the 1996 RAE that there was sufficient data for a full analysis to be undertaken.

Our analysis focuses on a comparison of the mean of all TQAs, for all subjects across each institution, as published at January 2000, with the mean RAE scores from the 1996 exercise, representing the quality of research activity over a four-year period from January 1992 to December 1995.

To verify the relative contribution of research criteria on teaching quality assessment outcomes, the impact of reputational and resource variables, which may be influencing the TQA scores is explored. The principal data source, the annual *Times Higher Education Supplement* league tables, assigns higher education institutions scores and rankings under a number of headings, including:

- *teaching*: mean of all subject reviews across the institution by the funding bodies, Quality Assurance Agency and Ofsted (the education ‘watchdog’) and published on the QAA website at January 2000 [later referred to as TQA];
- *research*: mean research assessment exercise score per staff member, based on the 1996 exercise on a 1–7 scale [later referred to as RAE];
- *entry standards*: mean average A-level point scores on entry of first-year, first-degree students [later referred to as SES];
- *student/staff ratio*: mean ratio for full-time equivalent (FTE) total student numbers on non-franchised courses and total teaching only, plus teaching / research staff [later referred to as SSR];
- *library and computer spending*: spending averaged over three years on the academic services; central libraries and information services and central computers and computer networks

per FTE student numbers. Includes expenditure on all libraries, learning resource centres and non-administrative computing [later referred to as LCS];

In addition, in this analysis, a variable is included that describes the ‘age-type’ of the university classified into four age groups: ‘ancient’, ‘pre-1945’, ‘modern’ and ‘post-1992’.

Exploratory analysis of TQA and RAE scores

The exploratory analysis examines the relationship between the mean TQA scores for the period 1993–2000, as dependent variable, and the above-mentioned independent variables.

Mean TQA scores for the four types of universities display small but clearly discernible differences. Over the period 1993–2000, the ‘ancient’ universities achieve the highest TQA scores, followed by the ‘pre-1945’ category (Figure 1).

There is an even more pronounced clustering of RAE scores, based on the age-based reputational hierarchy. The ‘ancient’ universities achieve, on average, by far the highest RAE scores, followed again by the ‘pre-1945’ institutions (Figure 2).

In line with the previous results, the variable ‘student entry standards’ closely follows the hierarchy classification, with drastic differences existing between the ‘post-1992’ institutions and the other three groupings (Figure 3). While student entry standards are but one factor which influence a student’s choice of university, they are a feature which can be viewed as self-perpetuating. In other words, universities that are perceived as being more difficult to gain entry to, attract more applications from students with the highest grades, and further enhance their academic reputations.

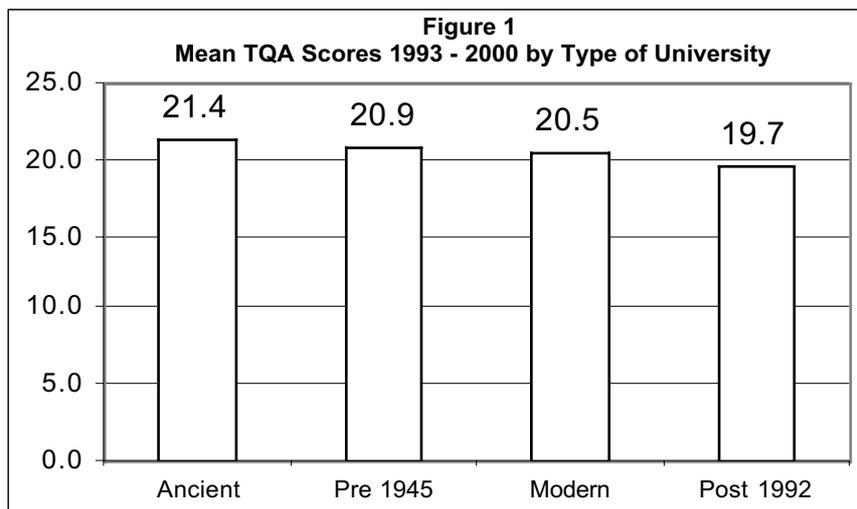


Figure 1: Mean TQA scores, 1993–2000 by ‘age-type’ of university

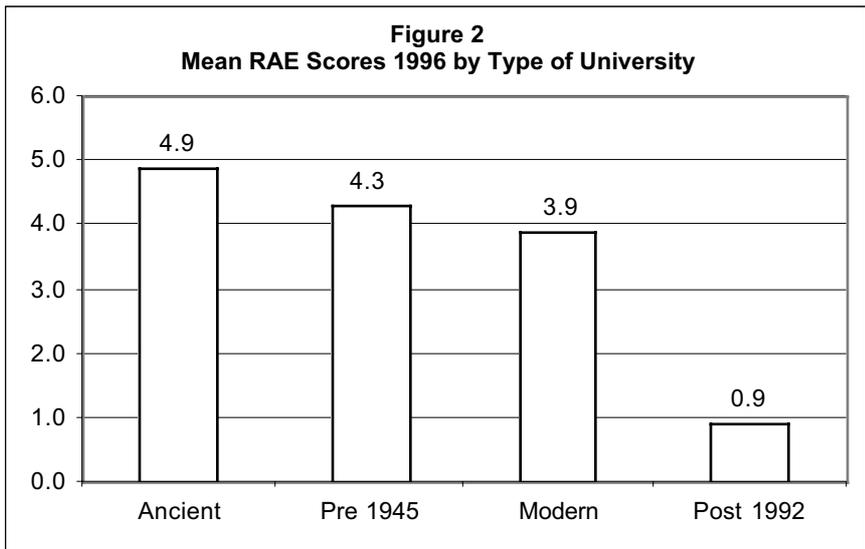


Figure 2: Mean RAE scores, 1996 by ‘age-type’ of university

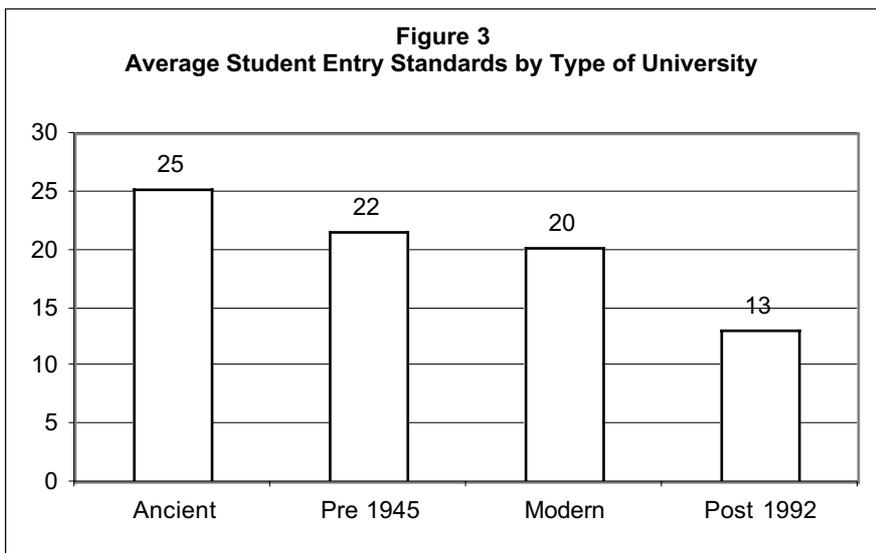


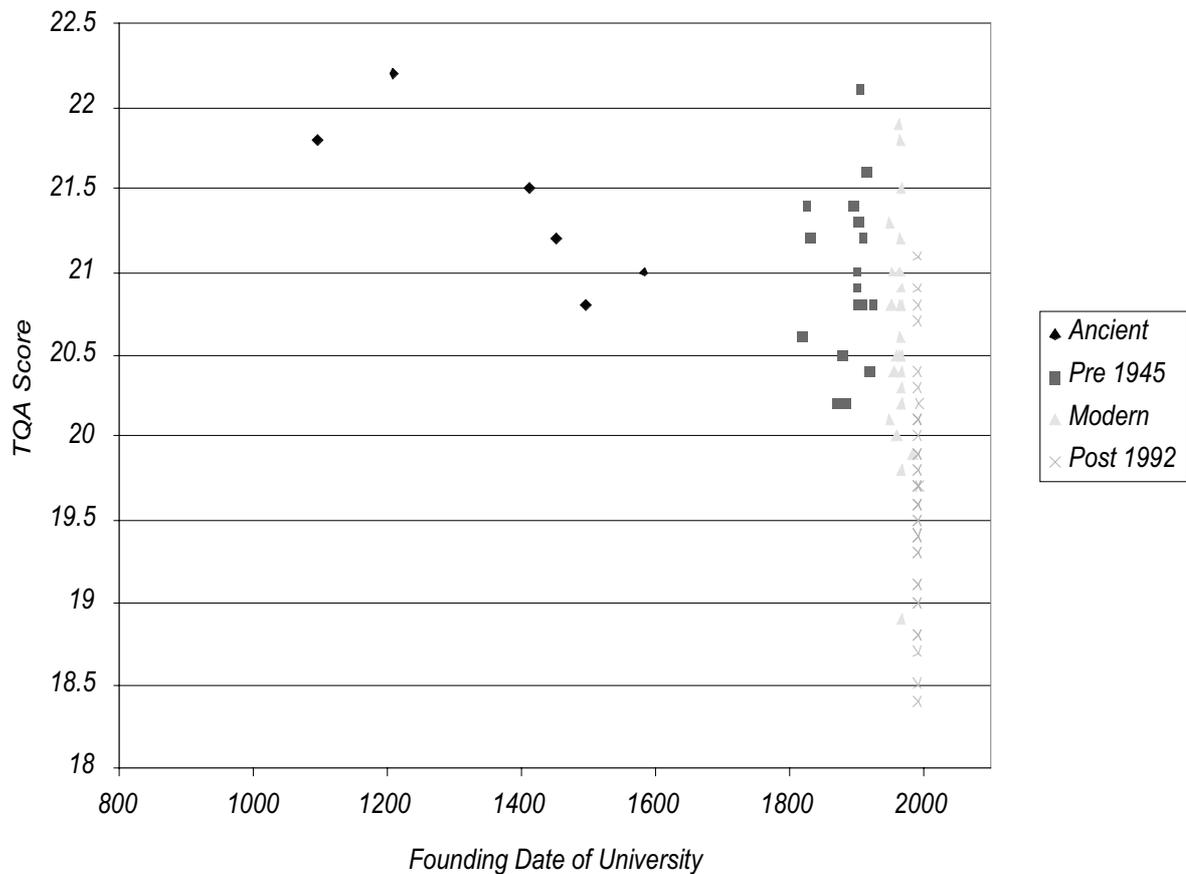
Figure 3: Mean student entry standards, 1998-2000 by ‘age-type’ of university

Figure 4 depicts a scatterplot of the age of the university and its TQA scores. Except for two of the ‘ancient’ universities (Cambridge and Oxford), there is little difference in the TQA scores achieved by these universities and their ‘pre-1945’ counterparts. ‘Pre-1945’ universities, in turn, achieve scores similar to those of ‘modern’ universities. Ignoring two of the bottom ‘modern’ outliers, more than half of the ‘post 1992’ universities, however, score lower than any of the other university groups in the scatterplot.

Figure 4: Mean TQA score by ‘age-type’ of university

A much more pronounced dualisation of universities is revealed by the scatterplot of mean TQA and RAE scores. In Figure 5, two distinct clusters can be identified, one including exclusively ‘new’ universities with a weighted average RAE score below 2, and a second with universities

Figure 4
Comparison of Mean TQA and Age of University



from the other three groups, with average RAE scores from 2 to 7. The vertical clustering around the TQA axis, meanwhile, is less pronounced, with many 'post-1992', new universities achieving scores comparable to older institutions.

What is important about this relationship is that it indicates the presence of strong interactions among a number of variables, which in turn affect TQA scores; a point noted by Yorke (1997) in connection with his identification of a strong collinearity amongst the component scores of league-tables.

The presence of these interactions between the variables is confirmed by the correlation matrix depicted in Table 1. This matrix displays zero order (bivariate) correlation coefficients for each possible pairing of the dependent and independent variables. While all correlation coefficients exceed 0.500, it is notable that the strongest interactions can be detected between the RAE score and student entry standards [SES], followed by those between TQA and SES, and TQA and RAE.

Figure 5
Comparison of Mean TQA and RAE Scores by Type of University

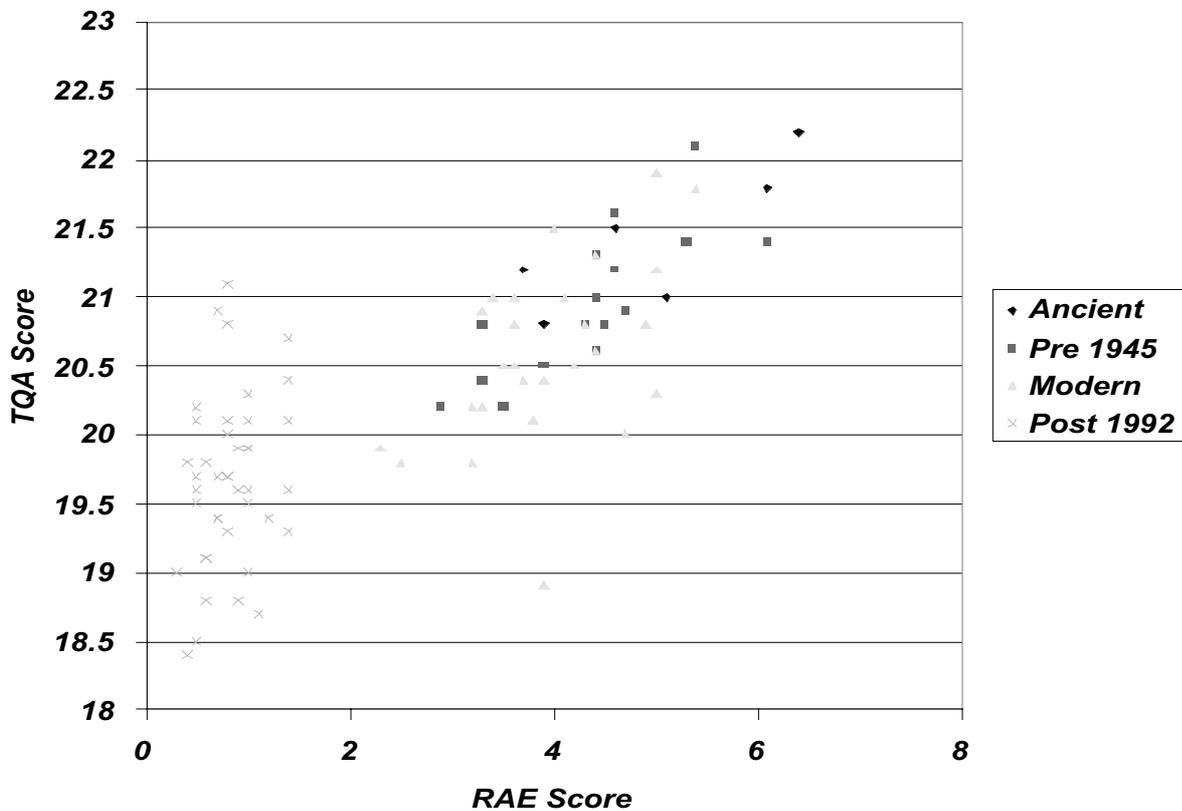


Figure 5: Mean TQA and RAE scores by ‘age-type’ of university

Table 1: Zero Order Correlation Matrix

	RAE	SES	SSR	LCS
TQA	.746	.763	-.653	.606
RAE		.904	-.666	.675
SES			-.672	.606
SSR				-.561

The presence of strong multi-collinear relationships, driven by SES and RAE is further confirmed by the varimax rotated, principal component loadings derived from this data set (Table 2). These loadings reflect the bearing an individual variable has on an artificially-generated composite component, which itself seeks to explain as much of the variation in the whole data set as possible.

Table 2: Varimax Rotated Principal Component Matrix

1 st COMPONENT	Loadings	Eigenvalue 3.754	Explained Variance 75.08%
TQA	.410		
RAE	.823		
SES	.838		
SSR	-.323		
LCS	.298		

The combination of the high loadings on RAE and SES, together with the fact that the first component explains over 75% of the variation, indicates that the variation within this data set is driven by a component which is strongly affected by those two variables, and only weakly by the contributions made by other variables (notably TQA, SSR and LCS). In other words, much of the variation of the variables in this data set can be attributed to the two variables RAE and SES.

In order to identify which variables best predict a university's TQA score, TQA scores (as dependent variable) were regressed with the three independent variables. Due to the likelihood of high multi-collinearity between SES and RAE, this analysis examines two multiple regression models, one including SES [and excluding RAE] (Table 3), and one including RAE [and excluding SES] (Table 4). In the first model, as concerns the size of standardised coefficients, student entry standards [SES] appear to make the strongest contribution to the equation, followed by the student-staff ratio [SSR]. The latter has, as expected, a negative coefficient (i.e. higher scoring universities had smaller student staff ratios).

Table 3: Multiple Regression of TQA with SES, SSR and LCS

Dependent Variable	TQA		
Independent Variables	SES	SSR	LCS
Coefficient			
Unstandardised	.086	-.058	.000
Standardised	.517	-.020	.178
Sign. Probability	.000	.021	.031
R Square	.637		
Sign. Probability F	.000		

These results are closely mirrored by the second model, which regresses TQA with RAE scores, the student-staff ratio [SSR] and library and computing spend [LCS]. Judging again by the standardised coefficient, in this model, the RAE score appears to be the strongest predictor of TQA scores, followed by the much weaker SSR.

Table 4: Multiple Regression of TQA with RAE, SSR and LCS

Dependent Variable	TQA		
Independent Variables	RAE	SSR	LCS
Coefficient			
Unstandardised	.227	-.072	.000
Standardised	.486	-.253	.137
Sign. Probability	.005	.000	.130
R Square	.610		
Sign. Probability F	.000		

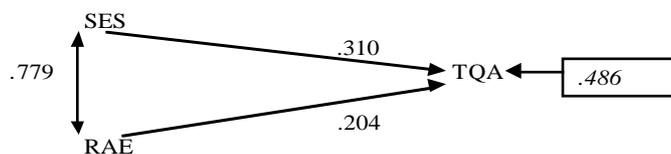
The primacy of SES over RAE as a predictor of TQA scores can be confirmed by applying a stepwise entry procedure to a regression of TQA against the entire variable set, using for example the typical default setting of probability of F entry at 0.05 and removal at 0.10. When applying this approach SES enters first, followed by SSR and LCS, with RAE being excluded on account of its high multi-collinearity with SES (producing results identical to those in Table 3).

A more detailed picture of the interaction between these variables can be gained by applying a partial correlation analysis. Focusing initially on the three variables, TQA, SES and RAE, this analysis highlights both the strength of the influence of SES and RAE on TQA and the strength of the interaction between these variables (Table 5a). Indeed, when depicting the relationship between the three variables in a conventional path diagram (Table 5b), SES and RAE together explain over half of the variation of TQA (with near equal strength).

Table 5a: Partial Correlation Analysis, TQA, SES, RAE

TQA with	SES		RAE	
	0 order	1 st order/RAE	0 order	1 st order/SES
	.763	.310	.746	.204
RAE with	SES			
	0 order	1 st order/TQA		
	.904	.779		

Table 5b: Corresponding Path Diagram



This analysis can be further expanded to control for all variables included in the data set. As shown in Table 6, higher order partial correlation coefficients confirm the strength of the relationship between TQA and SES, as well as, to a lesser degree, between TQA and SSR. Surprisingly, however, the higher order correlation between TQA and RAE drops to less than 0.10, indicating that RAE scores themselves correlate with other, in this case, resource-related variables.

Table 6: Higher Order Partial Correlation Analysis, TQA, SES, RAE, SSR, LCS

TQA with		0 order	3 rd order (all Vars)
	RAE	.746	.091
	SES	.763	.274
	SSR	-.653	-.229
	LCS	.606	.179

Conclusion

The teaching quality assessment exercise was proposed to be mission-sensitive. In assessing the quality of teaching in higher education, the assessors were expected to do this in relation to the stated aims of the institution. This was intended to give the post-1992 universities, which were strongly vocational in orientation and whose mission was largely one of wider access, a better opportunity to compete on an even footing with the older, established universities, which were more involved in research activity. The funding councils' intention does not appear to have achieved the expected result.

Our analysis of data for the period from 1993 to 2000, indicates that the determining factors in relation to high TQA scores are age of institution and reputation, in terms of attractiveness to students with high entry qualifications, and in terms of peer-assessed research. These variables are, in turn, strongly related to other variables such as the staff- student ratio and library and computer spending. Since reputations are established over a long period of time and resources appear to be most easily attained by institutions with strong reputations, post-1992 universities appear to be at a considerable disadvantage.

On basis of the, albeit tentative, analysis of this paper, we would conclude that, rather than representing an independent indicator of teaching quality, external quality review, as evidenced by the UK TQA scores, has largely served to confirm existing hierarchies, which may be determined by factors other than teaching.

REFERENCES

- BARNETT, R., 1994)], 'Judgment day for the judges', *Times Higher Educational Supplement*, No. 1143 : Sep 30/94, p.viii.
- FRAZER, M., 1992, 'Quality assurance in higher education', in *Quality Assurance in Higher Education: Proceedings of an international conference, Hong Kong, 1991*, Craft, A., (Ed.) 1992 (London, The Falmer Press).
- MORRIS, A., 1995, 'Bring those accountable to book', *Times Higher Education Supplement*, May 5/95, p.14
- TIMES HIGHER EDUCATION SUPPLEMENT, 2000, *League Tables 2000*, Apr 14/00
- YORKE, M., 1997, 'A good league table guide?', *Quality Assurance in Education*, 5(2), pp.61–72

