Misleading 'quality' measures in Higher Education: problems from combining diverse indicators that include subjective ratings and academic performance and costs

Diana Eugenie Kornbrot

Abstract

Quality measures are often derived from weighted sums of diverse indicators, including ordinal Likert items. This procedure can be dangerously misleading because it takes no account of correlations among indicators. It also takes no account of whether indicators are input measures, e.g. prior achievement of incoming students, or outcome measures, e.g. proportion getting a good degrees or student satisfaction. UK Higher Education data for England and Wales 2004-05 were analyzed taking these issues into account. Multiple regression showed, unsurprisingly, that 'bright' students with high prior achievement did well on all outcome indicators. A good research rating was associated with good degree performance and high completion rates, but not destination (in work or further training), or student satisfaction. Vice chancellor salaries and academic staff pay were positively associated with good destination outcomes. Worryingly, higher vice chancellor pay was associated with *lower* student satisfaction. The implications for evaluating university quality are discussed.

Misleading 'quality' measures in Higher Education: problems from combining diverse indicators that include subjective ratings and academic performance and costs

The 'quality' of Higher Education Institutes (HEIs) is a major concern: for actual and potential students, for government, for industry and for the staff of the HEIs themselves. So how should 'quality' be measured? Obviously, HEIs are complex and have many goals and so have many potential quality dimensions. The simplest solution is to decide on some indicator for each dimension and then take a weighted sum of the indicators. This is the approach taken by Newspapers when constructing league tables, national or international. There are at several problems with this approach.

1. Some indicators are metric, e.g. school achievement, proportion getting a 'good' degree, or money spent on libraries. Other metrics are derived from ordinal (Likert) items by inappropriate averaging. Thus the National Student Survey(NSS) of final year undergraduates asks students to give the extent of their agreement with 22 statements on a scale from 1-5 (5 Definitely agree, 4 Mostly agree, 3 Neither agree nor disagree, 2 Mostly disagree, 1 Definitely disagree, N/A Not applicable). The NSS score is then a Likert scale formed by averaging the scores from the 22 items. This procedure makes the unstated, and almost certainly wrong, assumption that the difference between strongly agree and agree is the same as between neutral and disagree.
2. The final metric is arbitrary. So that the consequences of a difference of 10, or even 100, points out of a maximum possible score of 500 in terms of chance of a student getting a good job or a good degree is unfathomable.
3. The weightings are arbitrary, with different weightings giving different rank orders. Thus, the Guardian, Telegraph and Times give different rank orders for UK universities.
4. The indicators, unsurprisingly, tend to be correlated. So HEIs with students with good prior achievement also have high rates of good degrees. HEIs who take students that are more able thus receive extra points for good degrees even though those same students might have done just as well at other universities.
5. League tables at the institution level take no account of discipline. Since most indicators are discipline specific the final results are likely to be biased by discipline mix. HEIs with many courses in popular disciplines (high average school achievement), or easy disciplines (low drop out rates), will achieve high scores. Bad news for physics and statistics! For this reason, many newspaper league tables do indeed provide discipline specific ratings. However, problems 1 to 4 apply at the discipline level as well as the institution level.

HEI quality indicators are potentially of use to many different groups. Intending students and their parents and sponsors (e.g. governments, charities etc.) want to choose the 'best' university. Governments, other sponsors and the HEIs themselves want to know if improvement is taking place - both absolute and relative. This paper takes a deeper look at what might constitute quality for a University in terms of benefits to undergraduates. The data comes from the Higher Education Statistics Agency (HESA) for Academic Year 05-06, available via The Times Higher Education Supplement, and the Sunday Times.

The first step was to separate input measures, such as school achievement (A & AS level points in UK system), from output measures such as percent getting a good degree (1st or 2.1 in UK system). Then two approaches are taken. The first is unashamedly exploratory, namely principal components analysis. What indicators of quality 'go together'? Perhaps, surprisingly this question has rarely been asked. The second approach *models* which input indicators predict particular output indicators. From the point of view of the intending students, this enables choice of the best HEI available, given their own prior achievement. From the point of view of HEIs or government it enables assessment of performance given both indicators within their control (spending of various sorts) and without their control (prior achievement of incoming students). A separate investigation explores the extent to which different weighting systems can affect the rank order of different HEIs.

This preliminary study looks only at the institution level in order to explore the approaches. Any serious evaluation of HEI quality should be at the discipline level, as this is where decision making takes place, be it for students, for the HEI or the government. The following resources discuss the use of performance indicators and benchmarking to enhance Higher Education Performance (Bekhradnia & Thompson, 2003; Bruneau & Savage, 2002; DfES, 2003; HEFCE, 2006; HESA, 2006b; Magd & Curry, 2003; Pursglove & Simpson, 2004). In addition, the problems faced in evaluating HEI quality are very similar to those faced for other complex situations such as 'level' of economic development of countries. In this situation also, composites are made by weighted sums of disparate indicators, with little attention to input as opposed to output indicators. The role of natural resources in economic development might be argued to be somewhat analogous to discipline in higher education.

The first task is to describe the HEI data set. Then the results of new analyses are described. Finally, the general implications for constructing quality measures are discussed.

Data Set for UK Universities 2005-2006

The data analyzed comes from the Higher Education Statistics Agency, (HESA, 2006a) which collects mandatory statistics from all UK Higher Education Institutions. HESA then sells CDs/books with summary details on students, staff, finance and destination. Cost is typically £50 per CD per Year. HESA makes some data available for free, but at the institution level or country level, but not the discipline level. HESA Performance Indicators available for free include, drop out rates by institution (but not discipline). In addition, the HOLIS page (http://www.hesa.ac.uk/acuk/maninfo/compareintro.htm) allows comparison of one chosen HEI by discipline and type of student for indicators such as drop out rate and class of degree with the *average* of selected other institutions. This is extraordinarily frustrating, as one cannot get a useful list of performance for all institutions in a given discipline, although one could get the information by making N/2 pairwise comparisons. Table 1 shows the indicators used in this study, obtained from Times Higher Education Supplement statistics summaries ("University Performance," 2007) that are obtained from HESA.

Table 1. Indicators of Higher Education Quality from Times Higher Education (THES) Statistics

Indicator	Code	Type	Mean	Min	Max
APoint : mean A&AS UCAS tariff points on entry	AP	Input	300	170	525
Student:Staff: ratio of students to staff (high is low resource)	SS	Input	18.5	8.4	33.6
NSS: Student satisfaction [a,b]	NSS	Input	14.9	13.7	16.1
RAE: Research Assessment Exercise	RAE	Input	3.7	0.5	6.6
Good Hons: Good degree: % with 1st or 2.1	GH	Output	59	39	88
Destination: % in work or advanced education	Dest	Output	63	42	87
Complete: % completing degree	Com	Output	84	67	99
£LibComp: Library & computer spend: £1000s: 3 year average	LM	Input	599	307	1656
£Facility: Facilities spend: £1000s: 3 year average	FM	Input	222	57	487
VCpay: Vice chancellor's pay: £1000s	VC	Input	165.5	104.3	305.0
ACpay: Mean full time pay: £1000s	ACP	Input	36.8	21.7	43.2

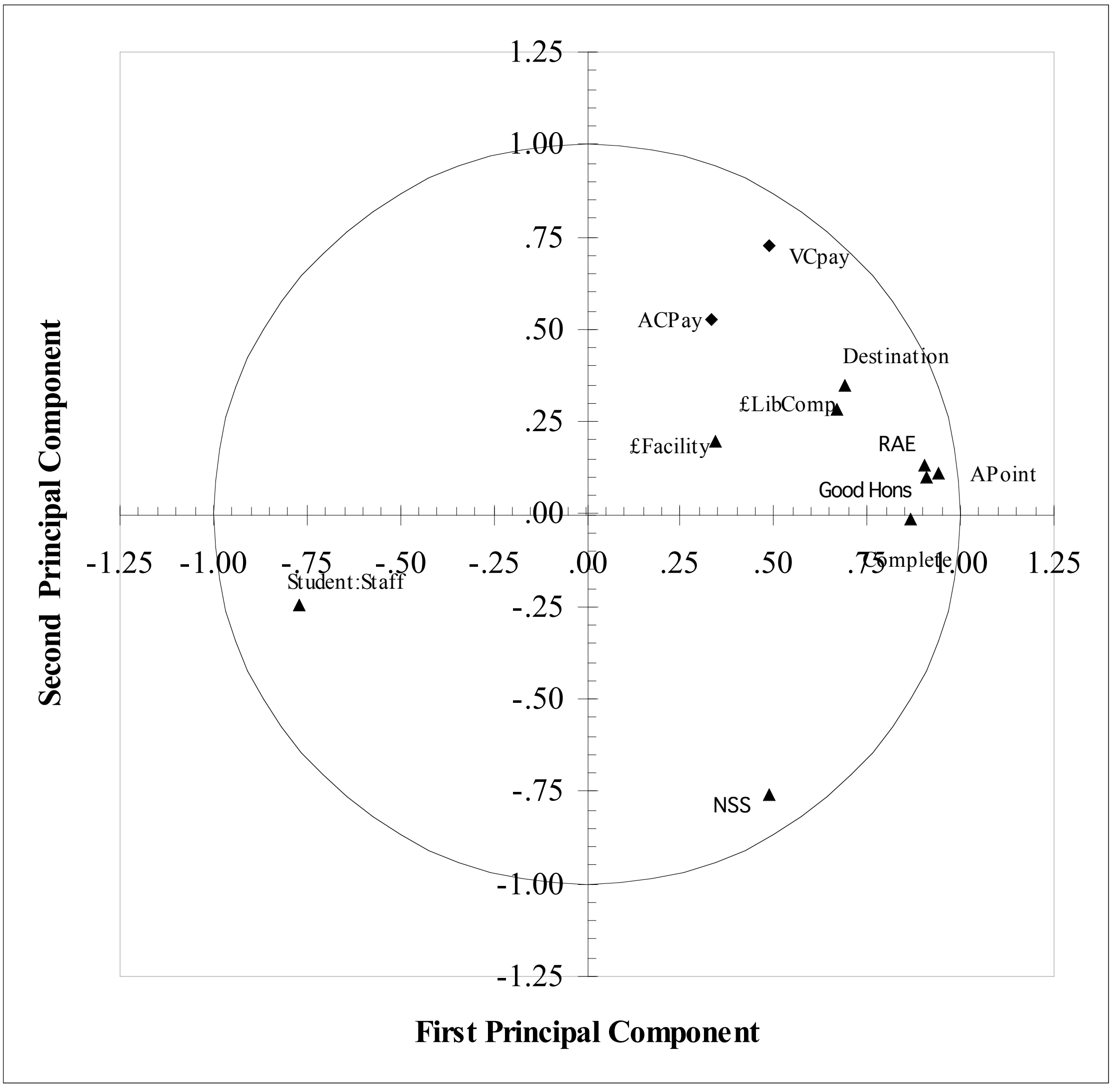
Note a: The NSS score ised bu THES is the mean of the scores on: teaching; assessment and feedback; academic support; organisation and management; and learning resources.
Note b: Universities not included, as no data: Cambridge, City, E. London, Oxford, South Bank, Warwick. All Scottish.
Note c: Universities excluded from because average academic pay was so low as to suggest a mis-recording: Lancaster (£21,686 Napier (£25,486 down 26% on previous year).

Analysis

Principal Components Analysis

A principal components analysis was conducted on the correlation matrix for 107 institutions, for all the variables in Table 1[1,2]. There were two components with eigenvalues greater than 1. The component loadings after varimax rotation are shown in Figure 1[3]. (The appendix shows the correlation matrix and component loadings)

Figure 1. Varimax rotated component loadings for the first two principal components



The first component of the varimax solution, accounting for 50.0% of variance, loads more than 0.45 on the output variables: probability of a good degree (Good Hons), probability of completion (Complete), and probability of being in employment or further training soon after graduation (Destination), and, to a lesser extent, on the subjective outcome indicator, student satisfaction (NSS). This component also loads on the input variables: starting academic achievement (APoints), research rating (RAE), vice chancellor pay, with a negative loading on student/staff ratio (fewer students per academic is better[4]). The second component, accounting for 15.6% of variance, loads *negatively* on the student satisfaction output variable, but positively on the input variables of vice chancellor pay (strongly) and other academic pay (less strongly). Thus, after allowing for the rotated first component, student satisfaction appears to be negatively related to staff salaries.

Modelling Student Outcomes

Four student outcome measures were modelled using multiple regression. These were % good degrees (Good Hons), % completion (Complete), % in work or advanced eduaction (Destination), and student satisfaction (NSS). All measures were normalized to z scores in the usual manner by subtracting the mean and dividing by the standard deviation. All regressions were performed on these z-scores. Since the predictor variables are correlated, a forward stepwise regression was also conducted. The final models presented, for each dependent variable, include all predictors with a significance level lower than 0.10. Once the non-significant predictors had been excluded, the included predictors were all statistically significant at the 5% level[5].

Table 2. Multiple regression parameters for outcomes with inputs as predictors

Predictor	Code	Good Hons Adjusted r^2 =.818				Complete Adjusted r^2=.753			
		B	se B	p null	ES	B	se B	p null	ES
APoints	AP	.744	.084	<.0005	.431	.790	.106	<.0005	.355
RAE	RAE	.182	.084	.032	.043	.216	.099	.031	.045
£LibComp	LM					-.190	.063	.003	.082
VCpay	VC								
ACpay	ACP								

Predictor	Code	Destination Adjusted r^2=.535				Satisfaction (NSS) Adjusted r^2=.404			
		B	se B	p null	ES	B	se B	p null	ES
APoints	AP	.300	.093	.002	.097	.769	.108	<.0005	.380
RAE	RAE								
£LibComp	LM	.218	.092	.019	.056				
VCpay	VC	.224	.087	.012	.065	-.608	.094	<.0005	.337
ACpay	ACP	.184	.074	.015	.060				

Note. ES is effect size as measured by partial eta squared.

Table 2 summarizes the models for each all four dependent variables. For each predictor, it shows the regression coefficient, B, standard error (se B), the p-value under the null hypothesis (p null), and the effect size partial eta squared (ES). There were 101 Institutions for the first three dependent variables and 93 for student satisfaction (excluding Scottish and four English).

Table 2 tells an interesting story. Unsurprisingly, student prior achievement, in terms of A levels UCAS tariff points on entry, is a strong predictor of *all* outcome measures. Conversely, the predictors student:staff ratio and facilities spend did not approach statistical significance for *any* dependent variable (lowest p-value =.266). There was a small effect of research excellence over and above A-level points for university performance in terms of both good degrees and completion. Library and computing spend had a negative effect on completion (after allowing for Apoint and RAE) and a positive effect on destination (after allowing for APoint, VCpay and ACpay), but no effect on good degrees. A high proportion going to a destination of work or further training was associated with higher pay for both vice-chancellors and ordinary mortals. The most surprising finding was that student satisfaction was *negatively* correlated with vice chancellor pay.

Creating League Tables

League tables can be constructed using combinations of any or all of the measures in Table 1. Since there is substantial correlation between measures, an obvious question is, 'Does the combination algorithm make much difference?'. In an attempt to answer this question, the 'top 10' universities have been constructed in several different ways. The published list produced by the Sunday Times for 2006 ("The Sunday Times university league table 2006," 2006) serves as a comparison. This list uses the following 9 measures (weightings in brackets): teaching quality (100); NSS (150); Heads rating (100); RAE (200); A-level points (250); % employment/further training (100); % good honours (100); staff:student ratio (100); and an arbitrary correction for drop out. As an illustration, four further rankings were created from the measures in Table1. Input is the average ranking of 3 indicators: A-level points; staff per student; and RAE. Output is the average ranking of 4 indicators: % good honours; % completion; % employment/further training; and NSS. Input & output is the average ranking of all 7 indicators; and NSS is just the ranking of NSS alone. Only the 82 universities for which there are no missing data are included (not Cambridge, Oxford, Warwick and Scottish institutions). Table 3 shows the rankings according to all these systems, for institutions that were in the top 10 for at least one ranking system, and number of systems for which they are in the top 10 (# top). Institutions are shown in bold for rank systems where they are in the top 10 and in italics for rank systems where they are *not* in the top 25. Clearly, the ranking system matters. Only York is in the top 10 for all measures. NSS produces markedly different rankings than other systems. This is as expected since NSS has relatively low correlations with the other indicators (see the Appendix).

Table 3a. Top 10 universities according to Sunday Times, Input & Output Input (3)

Sunday Times (9)		Input & Output (7)		Input (3)	
4	**York**	**6**	**York**	**4**	**York**
6	**Bristol**	**3**	**Bristol**	**6**	**Bristol**
9	**KCL**	**5**	**KCL**	**5**	**KCL**
1	**LSE**	**1**	**LSE**	**2**	**LSE**
3	**UCL**	**2**	**UCL**	**3**	**UCL**
7	**Bath**	**7**	**Bath**	14	Bath
16	Cardiff	**8**	**Cardiff**	**10**	**Cardiff**
2	**Imperial**	**4**	**Imperial**	**1**	**Imperial**
8	**Nottingham**	**10**	**Nottingham**	13	Nottingham
5	**Durham**	12	Durham	*27*	*Durham*
10	**Manchester**	14	Manchester	**7**	**Manchester**
12	Loughborough	18	Loughborough	*30*	*Loughborough*
18	Royal Holloway	**9**	**Royal Holloway**	12	Royal Holloway
35	*Aberystwyth*	*38*	*Aberystwyth*	*41*	*Aberystwyth*
42	*Bath Spa*	*44*	*Bath Spa*	*67*	*Bath Spa*
72	*Chester*	*54*	*Chester*	*64*	*Chester*
43	*Chichester*	*46*	*Chichester*	*72*	*Chichester*
31	*East Anglia*	*26*	*East Anglia*	17	East Anglia
39	*Lampeter*	*48*	*Lampeter*	*52*	*Lampeter*
17	Leicester	15	Leicester	*28*	*Leicester*
11	Southampton	13	Southampton	**9**	**Southampton**
20	Sussex	23	Sussex	**8**	**Sussex**

Table 3b. Top 10 universities according to Output (4), NSS(1) and according to number of inclusions in top 10 rankings

Output (4)		**NSS (1)**		**# top**
8	**York**	**7**	**York**	5
1	**Bristol**	22	Bristol	4
6	**KCL**	*29*	*KLC*	4
3	**LSE**	*28*	*LSE*	4
5	**UCL**	15	UCL	4
2	**Bath**	*31*	*Bath*	3
7	**Cardiff**	23	Cardiff	3
13	Imperial	*67*	*Imperial*	3
10	**Nottingham**	*36*	*Nottingham*	3
4	**Durham**	11	Durham	2
23	Manchester	*58*	*Manchester*	2
9	**Loughborough**	**1**	**Loughborough**	2
12	Royal Hollwy	**8**	**Royal Hollwy**	2
36	*Aberystwyth*	**9**	**Aberystwyth**	1
38	*Bath Spa*	**10**	**Bath Spa**	1
53	*Chester*	**6**	**Chester**	1
37	*Chichester*	**5**	**Chichester**	1
27	*East Anglia*	**4**	**East Anglia**	1
46	*Lampeter*	**2**	**Lampeter**	1
11	Leicester	**3**	**Leicester**	1
16	Southampton	*38*	*Southampton*	1
32	*Sussex*	*51*	*Sussex*	1

Another way of examining the difference between ratings is to look at the range between an institution's best and worst rankings. For the 3 input indicators the mean range was 18, with a maximum of 63. For the 4 output indicators (including NSS) the mean range was 32, with a maximum of 70. League tables are clearly a creative science.

Discussion

These analyses show features of HEI quality that are not, and cannot be, obtained from standard weighted addition measures of quality. They are at the institution rather than the institution within discipline level. However, similar problems will occur at this more detailed level. In particular, the finding that bright students do well wherever they are is likely to be duplicated across all disciplines, see also Bekhradnia & Thompson (2003). Nevertheless, at the discipline level other factors more under an institution's control (research, pay, student to staff ratios, facilities spend) may also play a part. Such effects may well be masked by the large discipline specific effect in all the student outcome measures. Effects other than student prior achievement are likely to be affected by discipline mix. For example, the association between academic pay and good destination rates may be mediated by a large number of clinical courses that both have high staff pay and almost all students proceeding to employment or further training. It is far from clear why student satisfaction should be so strongly *negatively* linked with vice-chancellor pay, but the finding certainly gives pause for thought.

Although, for illustrative purposes, I have included Scottish universities in the analyses that do not include NSS, it might be that differences in the education system and traditions and in the available data would make it more appropriate to analyze them separately.

In practice students, governments and HEIs make choices or take actions mostly at the discipline level. Consequently, the main future application of these methods will be at that level. The institution level analyses are presented simply as an example of what can be achieved when inputs are distinguished from outputs, and correlations are properly considered.

In summary, the methods presented here expose interesting features about the quality of undergraduate education at the institution level. They have even greater potential for a deeper understanding at the, far more important, discipline level.

References

Bekhradnia, B., & Thompson, J. (2003). Who does best at university? [Electronic Version]. Retrieved 8 May 2007 from http://www.hefce.ac.uk/learning/whodoes/.

Bruneau, W., & Savage, D. C. (2002). *Counting Out the Scholars: The Case Against Performance Indicators in Higher Education.* Toronto: James Lorimer & Company, Ltd.

DfES. (2003). *The Future of Higher Education,* . Norwich: Department for Education and Skills.

HEFCE. (2006). Review of performance indicators: Consultation to inform the review [Electronic Version]. Retrieved 8 May 2007 from http://www.hefce.ac.uk/pubs/hefce/2006/06_34/.

HESA. (2006a). Statistics [Electronic Version]. Retrieved 8 May 2007 from http://www.hesa.ac.uk/.

HESA. (2006b). *Students in Higher Education 04-05*: Higher Education Statistics Agency, http://www.hesa.ac.uk/products/pubs/home.htm.

Magd, H., & Curry, A. (2003). Benchmarking: achieving best value in public-sector organisations. *Benchmarking: An International Journal, 10* (3), 261-286.

Pursglove, J. D., & Simpson, M. (2004). Longitudinal trends in university financial performance. *International Journal of Business Performance Management, 6* (1), 1-21.

The Sunday Times university league table 2006 [Electronic (2006). Version]. *Timesonline* Retrieved 8 May 2007 from http://extras.timesonline.co.uk/stug2006/stug2006.pdf, http://www.timesonline.co.uk/tol/life_and_style/education/sunday_times_university_guide/article1324495.ece.

University Performance [Electronic (2007). Version]. *The Times Higher Education Supplement: statistics*. Retrieved 8 May 2007 from http://www.thes.co.uk/statistics/university_performance/league_tables/2006/main.aspx.

Notes

1. There was some non-normality in the raw variables So, as a further check, identical analyses were performed on data with all variables, except student:staff ratio, log transformed. (Log student:staff ratio was clearly non-normal whereas the raw scale was not reliably different from normal.) The loading patterns were effectively identical to those for the raw variables. Missing values were omitted pairwise, so the correlations, where NSS is not one of the variables, include Scottish universities, Oxford, and Cambridge.
2. A similar analysis was performed omitting missing values listwise, and thus excluding these universities. The results for the listwise and pairwise deletions gave the same loading patterns. The numbers reported are for pairwise deletions as this includes many more universities.
3. Oblique rotation, oblimin, gave axes with correlation of 0.057, and hence was not further considered.
4. Staff per student might seem the more natural measure, as a high value should be 'good' for students. However, students per staff (a number >1) is always quoted by convention and this variable does have an approximately normal distribution. For this reason this paper maintains the convention, so student:staff is negatively correlated with the other variables.
5. Analysis of the residuals showed no variable with points that had Cook's leverage greater than 2. The residuals did show some small trends of an increases as a function of the dependent variable. Consequently the raw data were log transformed and then standardised. Unfortunately, this resulted in 4 universities having

Cook's values higher than 2 in at least one analysis. It did not remove the trend in some residuals. The results with the standardised log variables were similar, except that some RAE effects that were significant in the unlogged data failed to reach significance in the log transformed data.

Appendix : Correlation Matrices and Component Loadings

Table A1

Correlations for the raw data with pairwise deletion
Where codes are as used in table

	AP	SS	NSS	RAE	LM	FM	GH	Com	Dest	VC	ACP
AP	1.000	-.664	.313	.869	.635	.284	.902	.851	.634	.562	.332
SS	-.664	1.000	-.195	-.696	-.601	-.294	-.683	-.560	-.619	-.469	-.379
NSS	.313	-.195	1.000	.311	.078	.028	.267	.374	.157	-.247	-.071
RAE	.869	-.696	.311	1.000	.558	.345	.829	.788	.570	.510	.449
LM	.635	-.601	.078	.558	1.000	.295	.640	.430	.583	.474	.154
FM	.284	-.294	.028	.345	.295	1.000	.296	.259	.136	.279	.092
GH	.902	-.683	.267	.829	.640	.296	1.000	.782	.615	.504	.255
Com	.851	-.560	.374	.788	.430	.259	.782	1.000	.497	.458	.343
Dest	.634	-.619	.157	.570	.583	.136	.615	.497	1.000	.567	.389
VC	.562	-.469	-.247	.510	.474	.279	.504	.458	.567	1.000	.421
ACP	.332	-.379	-.071	.449	.154	.092	.255	.343	.389	.421	1.000

Appendix : Correlation Matrices and Component Loadings

REJECT this version

Table A1

Correlations for the raw data with pairwise deletion

	APoint	Stud: Staff	NSS	RAE7	LiCo Money	FacM oney	Good Hons	Comp lete	Destin	VCpay05	ACPay05
APoint	1.000	-.664	.313	.869	.635	.284	.902	.851	.634	.562	.332
StudTOStaff	-.664	1.000	-.195	-.696	-.601	-.294	-.683	-.560	-.619	-.469	-.379
NSS	.313	-.195	1.000	.311	.078	.028	.267	.374	.157	-.247	-.071
RAE7	.869	-.696	.311	1.000	.558	.345	.829	.788	.570	.510	.449
LiCo Money	.635	-.601	.078	.558	1.000	.295	.640	.430	.583	.474	.154
FacM oney	.284	-.294	.028	.345	.295	1.000	.296	.259	.136	.279	.092
Good Hons	.902	-.683	.267	.829	.640	.296	1.000	.782	.615	.504	.255
Comp lete	.851	-.560	.374	.788	.430	.259	.782	1.000	.497	.458	.343
Destin	.634	-.619	.157	.570	.583	.136	.615	.497	1.000	.567	.389
VCpay05	.562	-.469	-.247	.510	.474	.279	.504	.458	.567	1.000	.421
ACPay05	.332	-.379	-.071	.449	.154	.092	.255	.343	.389	.421	1.000

Table A2.

Loadings after varimax rotation for the two components with Eigenvalues greater than 1.

Variable type	Variable	Component 1	Component 2
Input	A level Point (UCCA tariff) (AP)	.939	
	Student:staff ratio (SS)	-.768	
	Research, (RAE)	.905	
	Library and computing spend (LM)	.670	
	Facilities spend (FM)		
	VC pay (VC)	.488	.725
	Academic pay (AP)		.529
Output	Student satisfaction, (NSS)	.486	-.759
	% Good Degree (GH)	.911	
	% Completion (Comp)	.869	
	% Work or training (Dest)	.688	

Note. Loadings less than .45 (20% of variance) are shown as blank

Contact

Diana Kornbrot
University of Hertfordshire,
College Lane,
Hatfield, Herts,
AL10 9AB, UK.
d.e.kornbrot@herts.ac.uk.

Acknowledgement

Many thanks to the editors for much helpful advice.