



Using and Misusing Data in Higher Education

David Woodhouse

Commissioner for Development, CAA
Project Leader, CHEDS
UAE

May 2014

CAA Quality Series No. 7

Introduction

The **CAA Quality Series** comprises occasional publications, about two or three per year, on topics of interest to practitioners of quality assurance in higher education.

Specifically, the intent of the CAA Quality Series is:

- To contribute to the enhancement of quality practices in higher education in the UAE and more widely;
- To provide a means for sharing insights, research and analysis that is responsive to identified or emerging needs of those with responsibility for quality in higher education;
- To stimulate discussion and reflection on directions, evolution and progress in quality improvement relevant to UAE higher education;
- To provide contributions to the literature on quality assurance in UAE higher education that would otherwise not be available to a wide audience;
- To enhance public knowledge of QA, for agencies, for institutions and for the general public.

Contributions to the Series

Contributions, in Arabic or English, are invited from higher education quality assurance practitioners and educational leaders. The publications are expected to be scholarly and make a worthwhile contribution to thinking on or understanding of quality, addressing or responding to specific short-term policy issues as well as those of more general and longer-term relevance. They may be discussion papers, argue a particular case, or report the results of experiments or experiences. An indicative minimum word-length is 5000 words.

Anyone interested in contributing may contact the series editor, David Woodhouse, on david.woodhouse@mohesr.gov.ae.

The assistance of Ms. Reena Rajiv in managing the presentation and publication of the CAA Quality Series is gratefully acknowledged.

CAA Quality Series No. 7

ISSN 2227-4960

ISBN 978-9948-22-127-2

© 2013 Commission for Academic Accreditation

Ministry of Higher Education and Scientific Research

PO Box 45133

Abu Dhabi

United Arab Emirates

www.caa.ae

Contents

1.	Data	1
2.	Indicators	1
	2.1 <i>What are Indicators?</i>	1
	2.2 <i>Designing Indicators</i>	2
	2.3 <i>CHEDS</i>	3
	2.4 <i>Advantages and Disadvantages of Indicators</i>	4
3.	Learning Outcomes	5
	3.1 <i>Learning Without Outcomes!</i>	5
	3.2 <i>What is a LO?</i>	6
	3.3 <i>Problems with LOs</i>	6
4.	Academic Standards	8
	4.1 <i>Falling Standards?</i>	8
	4.2 <i>Assessing Standards</i>	9
5.	Benchmarking	11
	5.1 <i>What is Benchmarking?</i>	11
	5.2 <i>CHEDS</i>	12
	5.3 <i>AUQA</i>	12
6.	Ranking and Rating	13
	6.1 <i>What do Rankings Measure?</i>	13
	6.2 <i>Alternatives</i>	14
	6.3 <i>Actions by Academia</i>	15
	6.4 <i>Can you compare apples and oranges?</i>	16
7.	Conclusion	16
8.	References	17

I. Data

Historically, too much HEI planning has been data free. Writing in 1963, Sir Eric Ashby, former master of Clare College, Cambridge, said: "All over the country, these groups of scholars, who would not make a decision about the shape of a leaf or the derivation of a word...without painstakingly assembling the evidence, make decisions about admission policy, size of universities, staff-student ratios, content of courses and similar issues based on dubious assumptions, scrappy data and mere hunch." Since then, the wheel has turned, and HEIs gather enormous amounts of information and data – but use only a fraction of it.

When I was dean of a faculty in Australia in the 80s, I worked hard to encourage my staff to carry out student evaluations of teaching. 20 years later, as a head of the Australian quality agency, I was constantly trying to get institutions to reduce the number of evaluations they do because they had become at best useless and at worst counter-productive. Institutions could probably gather less data if they use it more effectively.

In most developed countries, there is now a great deal of HE data in the public domain, and institutions should use it, together with their own data, for comparative analyses, benchmarking and improvement. EQAs have a role in this as they constantly ask for data as evidence – i.e. information turned to a particular purpose – to show institutional performance and student achievements.

2. Indicators

2.1 What are Indicators?

An indicator can be an instrument or device that indicates the condition or performance of a machine or process. The instrument provides data, like the temperature of an engine. The meaning of the word indicator has been extended to refer to the data itself. In this sense, an indicator is a data item in context. So, for example, CHEDS (of which more in a moment) collects data and computes indicators. In this case, they are indicators of institutional performance. There are also indicators of the performance of an individual (e.g. a student or staff member) or of a whole educational system. Because indicators are so frequently about achievement (e.g. of goals or standards) or about performance, they are generally known as performance indicators (PIs). Indicators are like signposts.

Some common types of indicators are as follows:

- **Quantitative indicators** that can be presented as a number.
- **Qualitative indicators** that cannot (conveniently) be presented as a number.

- **Leading indicators** that predict the outcome of a process.
- **Lagging indicators** that report on what has happened.
- **Input indicators** that measure the amount of 'resources' used.
- **Process indicators** that represent the progress of a process (eg efficiency or productivity).
- **Output indicators** that reflect the outcome or results of a process.

2.2 *Designing Indicators* (Ewell & Jones, 1996)

PIs in higher education are particularly useful for three main purposes, namely to

- compare relative performances across activities, institutions or settings.
- monitor what is changing within a particular area, institution or context over time.
- examine the effects of intervention or policy change, between contexts or over time.

The purposes may be within the context of planning and/or QA/IE and/or benchmarking, and so on. To be useful, an indicator need not be causally related to what it is intended to reflect. For example, infant mortality rate is often used as an indicator of the overall health of a nation or group, but this overall health would not necessarily be materially improved by focusing all efforts on changing the value of this statistic.

A comprehensive set of indicators could usefully exhibit the following characteristics:

- **Leverage for action:** does the indicator provide concrete guidance for action?
- **Sturdiness:** the extent to which the numeric value of the indicator in question cannot be made to change without any real alteration in what it is designed to measure?
- **Credibility:** is the information content of an indicator likely to be considered trustworthy by a particular audience or user?
- **Ease of interpretation:** does the indicator convey a clear and consistent meaning to its intended audiences, both inside and outside the institution?
- **Perspective:** does the indicator relate to the points of view of its intended audience?

- **Comparators:** are metrics or standards available to chart progress through the indicator?
- **Technical adequacy:** is the indicator reliable and valid as a piece of data in itself, and how robust is it under typical conditions of biased or missing data?
- **Data availability:** is the indicator practically obtainable at a reasonable cost?
- **Materiality:** is the indicator valuable enough to justify the efforts needed to collect it?

2.3 CHEDS

Early in 2011, His Excellency Sheikh Nahyan Bin Mubarak Al Nahyan, the then Minister for Higher Education and Scientific Research, called for the creation of a national data collection and reporting function for higher education. The Data Warehouse within the Ministry had been collecting some data on the three federal institutions for several years, and the Commission for Academic Accreditation had been collecting some data on the institutions it licensed. However, what UAE needed was a single consistent higher education data base. The task was taken up by a small project group, and on 12 December 2011 the Minister issued decree No 347/2011 establishing the Center for Higher Education Data and Statistics (CHEDS). (www.cheds.ae)

The primary responsibility of CHEDS is to assemble a comprehensive data set of the higher education sector in the UAE, by gathering data from each higher education institution within the country. This data is then used to:

- produce a comprehensive annual report on higher education for the Minister;
- produce public reports on various aspects of higher education;
- respond to requests for data on the UAE higher education system; and
- facilitate international comparisons and national benchmarking.

To describe the UAE higher education system, CHEDS has identified 62 indicators spread across five functional areas:

- Institutional Information (13 indicators)
- Research and Innovation (8 indicators)
- Academic programs (21 indicators)
- Human Capital (13 indicators)
- Students & Learning Environment (7 indicators)

These indicators were carefully selected after considering both the data previously gathered in the UAE and international practice. Looking internationally gives reassurance that these indicators are widely seen as useful for their purpose, and assists the UAE in comparing itself internationally. Maintaining some continuity with current national practice facilitates longitudinal comparisons and trends. The consequence is that the Minister and other educational planners can have

institutional information that is recognized in the international research community as being indicative of quality higher education deployment. Assembling these indicators gives a comprehensive view of higher education in the country and allows for fact-based decisions, to enhance education in the UAE. CHEDS then defined 236 data items, across nine tables, to be collected from institutions, from which it calculates the value of these indicators.

CHEDS has now carried out four biannual data collections and the fifth is imminent. Each of these has been much more extended than we expected, as institutions have found it difficult to provide all the data required. Some institutions have mutually incompatible systems, some still operate manually, and some do not collect the data themselves. CHEDS has assisted by analysing each institution's submitted data and advising of gaps, and by visiting each institution.

UAE has a number of institutions (30+) in educational free zones and many of these have chosen not to respond to the data request. CHEDS' data does not give a full picture of the UAE HE sector, therefore, but it does cover over 90% of the student body.

Now CHEDS has this data, it has been able to relieve HEIs of some reporting tasks by reporting on their behalf. Most notable among these is the annual national reporting to UNESCO, where UNESCO complimented CHEDS on the quality and thoroughness of its report.

Also, now CHEDS is a part of the national system, it receives a large number of requests for information from many governmental and planning bodies, such as the Ministry of finance, the Prime Minister's Office, the President's Office, the Ministry of Foreign Affairs, Tawteen (the demographic planning body), and emirate governments. CHEDS is therefore clearly contributing to national development.

2.4 Advantages and Disadvantages of Indicators

(Ewell & Jones, 1996)

Well-designed and well-regarded indicators can help

- mobilize and align concerted action at all levels, from departments to the nation;
- institutions to communicate goals explicitly to students and the wider community; and
- support planning directed toward continuous improvement.

Drawbacks of indicator systems as management tools include the possibility of:

- focusing attention on data measurement, rather than data use: emphasising technical capabilities rather than human needs;
- failing to tell external stakeholders what they really want to know: institutions continue to focus on indicators of prestige over indicators of effectiveness; and

- creating false incentives for action: the real performance may be ignored in favour of influencing the indicators themselves (particularly if the stakes associated with 'poor performance' are high).

The third point is an example of Goodhart's law, namely that 'When a measure becomes a target, it ceases to be a good measure'. Examples of this are teaching to the test, chasing the indicators, and over-briefing interviewees for external reviews.

PIs should rarely be used singly or in isolation. Ratios are very valuable, and the most effective systems contain multiple measures that are mutually reinforcing. Indicator values should also be compared across different settings, at different times, or before and after intervention. In sum, PIs work best to inform an institution or system when there is a wide range of them and when they are clearly related to particular uses and users.

PIs should lead to questions rather than answers. More students passed this year: was it a brighter cohort, a better teacher, or softer grading? PIs are less well suited for rendering summative judgements about adequacy or performance.

Some people include the word 'quantitative' in the very definition of the term 'indicator', and want all indicators to be numeric, but we should allow for qualitative, descriptive, discursive indicators. Numbers give a sense of precision which may be false. An argument against non-numeric indicators is that they need interpretation – but that is true of numeric indicators also – it is just more obvious with non-numeric ones.

Measurement of a PI should be followed by investigations of what to do about any problems identified. In other words, PIs are unhelpful unless subjected to professional judgement and interpretation - which is what happens in an (internal or external) review. Such judgement should include consideration of the overall patterns that they suggest, rather than concentrating on small differences which may be the result of unimportant variations or simple chance.

3. Learning Outcomes

3.1 *Learning Without Outcomes!*

Early in my teaching career, my approach to teaching a course was straightforward. If I was asked to teach a course on linear algebra or compilers, I listed the topics I thought should be covered, told the students about those topics (it's called 'lecturing') and then examined whether they'd learned what I'd told them. If I'd been asked what outcomes I expected from the students, I'd have said, that they learn what I told them. Of course, I was already behind the times because Bloom's taxonomy much pre-dated my first lectures. But yet I don't feel too guilty, because my approach was common at that time. Since then, however, we have seen a revolution in the

approach to student learning, and now, of course, I'd be expected to think about that question in advance, and start by deciding in what ways I wanted the course to change the students, and what they'd be able to do at the end.

"I respect no study, and deem no study good, which results in money-making" (Seneca, Roman statesman and humorist). Or as my own DPhil supervisor (in abstract algebra) said: "If it is useful, it is not pure mathematics".

3.2 What is a LO?

A **learning outcome** (LO) may be defined as "what a learner is expected to know, understand and/or be able to demonstrate after completing a process of learning".

The LO spread to HE from the vocational sector. In the vocational sector, learning outcomes based on competencies are used to underpin the assessment of job-related skills. Once the notion of having to account for learning had been set in place, its extension to another part of the education system seemed natural, and the LO is a seductively simple concept. (Scott, 2009)

The Mastery learning movement (e.g. Bloom, 1981) proposed that the vast majority of learners are capable of achieving to the same extent, but that learners would take differing amount of time and input to achieve. Within mastery programs, learners must achieve specific LOs before being permitted to proceed to the next stage. This led in the 80s to Outcome Based Education (OBE) which puts emphasis on the outcomes of learning processes rather than the inputs or the processes themselves.

LOs are intended to give a clear indication to the learner of her or his destiny in the course. As a result, the proponents of OBE claim that is 'student-centred', in contrast to my early practice where the destination was revealed only by the teacher as the course proceeded. In fact, it is no more student-centred, as the student still depends upon the teacher to guide her/him along the road to the destination: knowing you are going to Muscat does not make you able to get there.

3.3 Problems with LOs

Not only are LOs deceptive in their apparent student-centredness, they are deceptive in terms of their precision. LOs "give the impression of precision only because we unconsciously interpret them against a prior understanding of what is required. ... The meaning of the evaluative terms used to specify the quality of knowledge, understanding or analyses is always relative to a context and so cannot be used to specify absolutes." (Hussey & Smith, 2002) The New Zealand and South African Qualifications Authorities tried to specify LOs with a definitive level of detail in their respective national qualifications frameworks, but whatever the level of detail that is used, either there are always words that need further definition in an endless

regression, or eventually we stop with an agreement on prior understanding of meaning.

“Outcome statements ... are open to very different interpretations. In trying to contain these differences, outcome developers make them more and more specific – but in the process, they get narrower and narrower, and also, longer and longer, and consequently more difficult for curriculum designers, teachers, and assessors, to work with. Yet, they never become transparent.” (Allais, 2009, Hall & Woodhouse, 1999) “The assumption that human capabilities can be unequivocally described and accurately communicated by means of language is unfounded. So, at best, written competency standards are rough and ready, though useful, guides, and ... it is not the words that are important but what they mean, and the extent to which what they mean is widely understood.” (Guthrie 2009) “The failure of the NQF in South Africa was inevitable, because of inherent flaws in the idea that specifications of learning outcomes in qualifications can increase the quality and quantity of educational provision” (Allais, 2013)

Scott (2009) gives the following example from a hypothetical competency-based carpentry course.

LO: After the period of learning the student will be able to: bang a nail into a plank of wood without splitting the wood.

At first glance, this seems like a straightforward learning outcome, but a carpenter might well ask, “which type of wood” or “which type of nail”. So it would be necessary to moderate the outcome so that it might become;

LO: After the period of learning the student will be able to: bang the appropriate nail into a plank from a range of commonly used timbers without splitting the wood.

But then the carpenter points out that there is no reference to accuracy or safety. So then we define the seeming obvious – but are confronted by another carpenter who notes that what is a common wood for some is not common for him; how was he meant to know what I meant or what his student was meant to learn? The only defence from the carpenter's demands is to either write with more and more specificity or greater generality. The problem with the former is that increased specificity starts to exclude many practices, while the problem with the latter is that writing very broad and general learning outcomes means that either no one is clear what the learning outcome is about or that you can work it out only if you have sufficient prior knowledge and understanding of the subject in question and its context. Hussey and Smith (ibid.) suggest that in order to explicate a phenomenon LOs must “parasitise” that which they are meant to be explaining.

Thus, the slavish attention to LOs and their measurement can be at best superficial and at worst counter-productive. When I came to the UAE three years ago, I was delighted to find that the CAA emphasizes LOs. I believe that this has been a positive emphasis. It has encouraged (forced?!) institutions to think more carefully about what they teach and how they assess it. However, this emphasis has led to some strange practices. Some institutions specify how much of each course objective is assessed by

each question on the final course exam, and then do some fancy acrobatic arithmetical manoeuvres to show the extent to which students have achieved each course objective, based on their performance on the exam. This sort of thing is reported to the CAA as success (or otherwise) in achieving the course's LOs. In my opinion, it is nonsensical. This opinion is supported by the fact that the students do not know to what extent they have achieved the course LOs. They just did the exam and tried to pass: this is the student equivalent of my first teaching where I told them the material and expected them to learn it. In neither case, are LOs involved.

I recall a senior UK colleague saying in a conference a few years ago 'when I hear the term LO I reach for my gun' (alluding of course to the quote that came from the Nazi regime in Germany). At the time I was surprised, but I have now come to see his point in the over-use and mis-use of LOs. It is difficult: "It is when the gods hate a man with unusual abhorrence that they drive him into the profession of a schoolmaster" (Seneca). (Of course, Seneca was Nero's tutor and was forced to commit suicide for alleged treason, so he had a particularly difficult row to hoe!)

4. Academic Standards

4.1 *Falling Standards?*

Here we have another area in which the over-use of apparently precise data or specifications can cause problems. One peril in trying to achieve common standards in a subject area across institutions is that this needs objectivity, which quickly becomes synonymous with quantification, while professional judgement is dismissed as impressionistic. However, a disciplinary "tradition depends on a community of scholars engaged in a common pursuit ... with some sense of what it means to engage in it well or badly. Standards are therefore inherent in the very notion of a discipline. Objective assessment becomes possible in the light of a growing familiarity with examples of good practice" and with the judgements that are made about them (Standish in Blake et al., 2013).

A concern has been expressed in many countries of falling academic standards, grade inflation, and pressure to pass students (especially fee-paying ones). Attention to this leads immediately to considerations of how to reliably assess students' performance and level of achievement – ie the academic standards attained. In the early 90s, the UK Secretary for Education, commenting on the work of the national quality agency (QAA), said (I paraphrase): "The QAA knows about the quality of HE, but does it know about the standards of HE?". This led to the Graduate Standards Programme (GSP), a project which lasted for a couple of years, and cost a couple of million pounds. The final report (QAA, 1996) said that "there is no general consensus, either within or outside higher education, that UK degrees are broadly comparable with each other in terms of equivalence of output standards"; and that "judgements of standards in many fields, including higher education, are ultimately rooted in the shared (and generally tacit) values of specialist communities. These values tend not to be

articulated explicitly, but are realised through the practice of that community. In consequence, the new entrant to a particular academic field tends to absorb them more or less unconsciously through participation in day-to-day activities and interaction with other academics. This may take place, for example, through debate...”.

However, while academics debate their research findings, fewer debate their marking and assessment practices, which can result in gross differences.

In the USA, similar comments are still being made: “By 2010, surveys by the Association of American Colleges and Universities (AAC&U) and NILOA found that more than three quarters of all institutions had developed outcomes statements to guide teaching and represent student learning. Proving far more difficult for institutions was moving from lofty, broad, and sometimes vague descriptions of student performance to demonstrable evidence that students had, in areas institutions deemed appropriate, indeed become proficient.” (Ewell, 2013). “In January 2011, the Lumina Foundation published its Degree Qualifications Profile (DQP) to challenge faculty and academic leaders in the U.S. to think deeply and concretely about aligning expectations for student learning outcomes across higher education. ... The Profile proposes sets of competencies in five areas of student learning (Specialized Knowledge, Broad Integrative Knowledge, Intellectual Skills, Applied Learning, Civic Learning) and addresses three degree levels (associate, bachelor, master)” (ibid).

4.2 Assessing Standards

In the UK, the QAA developed ‘subject benchmark statements’, as an attempt to bring more rigour and consistency to the judgement of whether a student had reached a certain level of performance. The statements were supposed to be precise enough to enable consistent judgments by different people or review teams; but not precise enough to become a fixed national curriculum. This was of course impossible, and what slipped was the former. They are not a national curriculum, but embedded in them are words (such as ‘appropriate’, ‘relevant’) that imply professional judgement, and mean that different people or review teams might come to different opinions. In the USA, the long-standing assessment movement has given rise to much advice on how to assess student performance (eg Rhodes, 2010).

Australia tried a similar process in more recent years. People noted that some statements in these subject specifications, if taken out of context, could refer to school, bachelor or master level. For example, the history specification set one LO as ‘Demonstrate knowledge of one or more periods of the past’. The authors seem to have had no understanding that a standard implies a level.

In the US, despite most institutions having developed outcomes statements to guide teaching and represent student learning, “it is proving far more difficult for institutions to move from lofty, broad and sometimes vague descriptions of student

performance to demonstrable evident that students had indeed become proficient” (Ewell, 2013).

In 2011 the Lumina Foundation published the Degree Qualifications Profile to assist in this. It would be familiar to us in the UAE in relation to the QFEmirates. It has 5 areas of competency, namely: Specialised knowledge, Broad integrative knowledge, Intellectual skills, Applied learning, Civic learning; at 3 levels: associate, bachelor, master

The above criticisms are not to say that the subject statements are without value. They certainly have helped with consistency, but the problem comes if they are expected to provide definitive incontrovertible conclusions. Professional judgement by those in the field still needs to be applied.

Furthermore, that judgement needs constant updating through academic interaction – which, as I noted above, is not sufficiently common. I have been a member of groups marking hundreds of end-of-school exam papers, and we have had discussion meetings to make our tacit knowledge and understandings explicit, and to (try to) ensure that we are forming the same judgements. At HE level, however, academics behave in a more individualistic fashion, and rarely go through such a process. Of course, it might not be easy: in a small department, there may not be anyone else with your depth of knowledge of your specific area, so such a discussion would need to involve several institutions.

D.R.Sadler has worked in this area for many years (beginning in Sadler, 1987) and has suggested that what is needed is for academics to agree on and share a judicious combination of ‘verbal descriptions’ and ‘exemplars’ of work that is typical of designated levels of competence or achievement.

Proposals to establish such ‘standards networks’ surface from time to time (eg James, McInnis & Devlin, 2002), but then subside because of the practicalities.

The UK and Denmark have external examiner systems, which help in this regard; there is increasing moderation of examinations and assignment; and some institutions use inter-institutional cross-marking – but there is still some way to go.

The difficulty is defining academic standards precisely enough that their meaning is clear to academics, students and the community, but not so tightly to suppress innovative ways of teaching and assessing students. This suggests that consideration of the learning-teaching process is important in the assessment of standards, despite attempts to remove process considerations to cater for the increasing variety of modes of learning and types of institutions. Processes include the seven ‘good education practices’ of Chickering and Gamson (1987) (on which the NSSE & AUSSE are based).

Thomas Jefferson, speaking on C18th US schools: “They commit their pupils to the theater of the world with just taste enough of learning to be alienated from

industrious pursuits, and not enough to do service in the ranks of science.” Even today, we sometime raise expectations but without providing the qualifications.

5. Benchmarking

5.1 What is Benchmarking?

The word **benchmark** means a point of comparison, usually with a connotation that it is a desirable level of performance. **Benchmarking** has come to mean a formal and structured process for carrying out such a comparison, and using the comparison to achieve improvement in one’s own performance. The word has been increasingly used in academia over the last couple of decades. Although the informal exchange of information has long been part of academic culture, benchmarking provides a formal and objective structure for this exchange. “Due to its reliance on hard data and research, benchmarking is especially suited for institutions of higher education in which these concepts are familiar. Benchmarking can help overcome resistance to change, provide a structure for external evaluation, and create new networks of communication between parts of the HEI.” (Alstete, 1995)

If I asked an institution in the first half of the noughties whether it was benchmarking, a typical answer was ‘we’re thinking about it’. In the second half of the decade, institutions said they were doing it, **but** mostly they were just getting data from other institutions and looking at it – little structured use was being made of it. At the other extreme is a formal and extensive benchmarking process, which might look as follows (Gott, 1999):

Planning

- Select a topic and appoint a benchmarking team
- Select benchmarking partner(s)

Collection

- Document own process and issues
- Identify indicators for measuring performance
- Collect data on current performance

Analysis

- Analyse and compare data; identify good practice and gaps
- Establish target goals for improved performance

Action

- Develop and implement an action plan to achieve the targets
- Monitor progress

By highlighting problem areas as well as the potential for improvement, benchmarking provides an incentive to change and assists in the setting of target

goals. Furthermore, its emphasis on understanding the processes underlying successful practice makes it a useful tool in establishing plans and strategies for achieving these goals. Benchmarking is thus a comprehensive and self-contained method for improving organisational practices: it highlights areas needing improvement, it provides objective data to illustrate the need for change in these areas, and it leads to the formulation of plans and initiatives for bringing about the required improvements.

5.2 CHEDS

In the UAE, CHEDS work is beginning to put out into the public domain data that will enable institutions to at least use that data for initial self-knowledge, and CHEDS further encourages institutions to identify self-selected benchmarking groups within which institutions will be willing to share the more detailed data they each have on themselves.

Some institutional data is quite sensitive, and some institutions have withheld specific data, particularly aspects of financing. Nonetheless, CHEDS has reported publicly by combining the institutions into clusters on four different dimensions (namely size, type, highest level of education, and location) and reporting aggregate values for each indicator for each cluster. Thus, even without communicating with another institution, each institution can compare itself to the averages of four different groups. Increasingly, institutions are telling us that they are using this data for benchmarking across the sector.

In addition to these CHEDS-determined clusters, an institution may request the aggregate values of specified indicators for a selected group of not fewer than five and not more than 10 institutions. These 'peer group reports' allow an institution to select institutions with which it considers itself comparable in some way on the specified indicators.

5.3 AUQA

The former Australian national quality agency (AUQA) carried out whole-of-institution audits. The federal Australian government collects a great deal of data on all HEIs and made this available to AUQA. When AUQA assembled an audit team to review one institution, the team would have access to data on that institution. But how would it evaluate that data? If the pass rate were 80%, the institution might be seen to be 'good', but with a pass rate of 40%, then 'bad'; similarly, a drop-out of 60% would be 'bad', etc. But would these judgements be valid?

For each audit team, AUQA did a cluster analysis on the whole higher education sector in relation to some of the most significant descriptors, and then provided the team with a comparison of the institution with the aggregate for 'similar' institutions. It might then be seen that all institution in the cluster were getting pass rates of 80-

90% (maybe they were the selective research-oriented institutions), so 80% was actually not very good. Or again, maybe all the institutions in the cluster were getting drop-outs around 50-60% - maybe they were the distance education specialists.

In other words, we compared like with like to come to fair conclusions. We also analysed trends to see whether year-to-year changes were significant or not, and investigated them in the former case – or, first, asked the institution whether it had investigated them!

6. Ranking and Rating

6.1 What do Rankings Measure?

One increasingly common mis-use of data is for ranking institutions. I call it 'mis-use' as largely the rankings are based on arbitrary indicators with arbitrary weightings and an emphasis on research or the vague notion of 'prestige'. Changing the weightings can change the order of the ranking by many positions. (For example, weight internationalisation highly in the Times ranking and MIT drops to about 300th in the world!) Yet they still fail to tell prospective students and their parents much about which institution to attend or why. Rankings are simplistic (Phil Baty).

The meaning of the word quality has shifted over the last 40 years. The word that best captured its meaning then is 'reputation', and it was applied to such things as Rolls Royce cars and Oxford and Harvard Universities. Since then, the word 'quality' has become 'democratised', and the phrase that best captures its current meaning is 'fitness for purpose' (where the various purposes may be manifold, complex, and even contradictory).

Rankings have emerged on this scene and purport to measure quality, whereas in fact they are mostly measuring (long-standing) reputation. In commenting on the THES and SJTU rankings, Marginson (2006) observed that they produce very different results "below the very top". Usher and Savino (2006) report that "Regardless of the ranking scheme employed, 'top universities' are almost always going to come out as top universities. The variation between rankings occurs lower down the scale; there, even small changes in methodology can change rankings significantly."

In 1906, a survey was done in the US to ascertain the reputation of US universities. In 1982, the survey was repeated. Only two institutions had changed in the top 14 in almost 80 years. Few other areas of activity (e.g. best-known businesses) would have changed so little in that time.

Persistence of Perceptions of Prestige

	1906	1982
1	Harvard	UC Berkeley
2	Columbia	Stanford
3	Chicago	Harvard
4	Cornell	Yale
5	Johns Hopkins	MIT
6	UC Berkeley	Princeton
7	Yale	Chicago
8	Michigan	UCLA
9	MIT	Michigan
10	Wisconsin-Madison	Wisconsin
11	Pennsylvania	Cornell
12	Stanford	Columbia
13	Princeton	Illinois-Urbana
14	Minnesota	Pennsylvania

Usher & Savino also say that “institutional ranking systems don’t measure what the authors think they are measuring. ... What our results here show is that most indicators are probably epiphenomena of some underlying ... feature that is not being measured. ... some ... ‘X factor’ ... Our guess is that age of institution, faculty size and per-student expenditure are probably excellent candidates to be these ‘X factors’.”

6.2 Alternatives

Attempts to avoid the worst effects of rankings include ratings, user-specified weightings, comparing only among similar institutions, and comparing at finer-grained detail.

Ratings

A defect of the linear ranking is that the difference between successive places is usually not statistically significant. The term ‘rating’ is sometimes used to denote a listing where the entities have been clustered in bands – eg quartiles, or quintiles as in the Australian ‘Good University Guide’. Of course, institutions at the boundaries can still be adversely (or positively) affected.

Weightings

The German CHE provides the basic data and allows users to specify their weightings. CHEDS has something like this on a future horizon.

Comparable Institutions

This is why CHEDS has reported publicly the aggregate performance on clusters of institutions which share a common characteristic. A broader example is the European U-Multirank. “One principle of U-Multirank is the comparability of institutions. In

rankings, institutions and programmes should only be compared when their purposes and activity profiles are sufficiently similar. [Exactly the approach taken by AUQA – above.] It makes no sense to compare the research performance of a major metropolitan research university with that of a remotely located University of Applied Science; or to compare the internationalisation achievements of a national humanities college whose major purpose is to develop and preserve its unique national language with an internationally orientated European university with branch campuses in Asia.” (CHERPA Network, 2011; see also www.umultirank.org)

Finer Detail

CHEDS’ 62 indicators span the range of academic activity. Within this, some indicators relate to research, some to teaching, some to supporting national development, and so on. (Obviously there is an overlap.) CHEDS is therefore producing a report in which the performance of each institution on each of these factors or aspects of education is represented on a spider graph, with the national quartile marks included for comparison.

6.3 Actions by Academia

One main problem with rankings, therefore, is that they are measuring something, but most of their readers think they are measuring something else. (They don’t come close to measuring what complex institutions do – Bob Morse.) It is essential that academia

- does not give them spurious validity by cheering whenever one’s own university scores highly,
- actively publishes well-argued criticism when their producers make inflated or misleading claims,
- help the public to distinguish between the information they often say they want and the information that their enquiries show that they actually need, and
- provides easily accessible information of the sort that the public needs. (Stella & Woodhouse, 2006)

People say ‘rankings are here to stay’ in the same breath as they complain about the rankings. Merely to accept them as a given is analogous to observing that ‘nuclear weapons are here to stay’, but then use this as a reason not to work for their containment and limitation. A large number of campaigns and reach-out strategies at various levels may be necessary for different stakeholders to help them use quality-related information in the most appropriate and helpful way. In both secondary and higher education institutions, quality literacy should become a part of academic and personal guidance and counseling. This would facilitate the emergence of a quality-literate society – a society that can take decisions informed by quality-related data. This is a priority area for quality assurance agencies the world over.

6.4 Can you compare apples and oranges?

It is odd that the stock phrase to warn against comparing the incomparable is 'you can't compare apples and oranges'. But why are apples and oranges innately incomparable? They are both fruit, aren't they? If we said 'you can't compare apples and motor cars' it might make sense. Or elephants and oranges. But surely one can compare different fruits ...

7. Conclusion

"Quality is never an accident: it is always the result of intelligent effort". (John Ruskin)
So, where is the evidence for quality? If there were a brief and simple answer to this question, it would probably have been found by now. In fact, it is so difficult that universities have even foregone funds. The following is from an Australian news report in May 2011: "**Universities have conceded** that the government was justified in its budget decision to defer \$95m of promised performance funding. The funds were deferred because **the sector has been unable to produce viable indicators of the quality of student experience and learning**".

This is a serious indictment of the modern HE sector.

Having said that, measuring teaching will always be harder than measuring research because of the following innate differences between the two activities:

Researcher A carries out research B; B is evaluated; this can be taken as a measure of A's research performance.

Teacher A teaches student B; B is evaluated; many influences other than teacher A have gone into student B's performance on the evaluation; therefore the evaluation cannot be taken as the TOTAL measure of A's teaching performance.

Surely the evidence must lie in a system of well-chosen indicators, set within a coherent model. The indicators must include input and process indicators, not only output ones. The indicators must relate to the concepts they are measuring, but must also relate to the audience for which the information is intended.

There is an increasing movement towards looking at process indicators that represent the extent to which students engage in the activities that predict desired learning outcomes (NSSE, AUSSE). These include such features as the seven 'good education practices' of Chickering and Gamson (1987) (on which the NSSE & AUSSE are based).

And institutional researchers are absolutely central to this process of building, presenting and using evidence.

However, no-one should be under any illusions that this will be easy.

8. References

- Allais, S.D. et al. (2009), 'Learning from the first qualifications Frameworks', *Employment Working Paper No. 45*. ILO
- Allais, S.D. (2013), 'National qualifications frameworks: a critical analysis of an important international approach to education reform', *ESRC Research Report PTA-026-27-2166*
- Alstete, J.W. (1995), 'Benchmarking in higher education', *ASHE-ERIC HE Report No. 5*, George Washington University, Washington DC
- Bender, B.E. & Schuh, J.H. (eds) (2002), 'Using benchmarking to inform practice in higher education', *New Directions for Higher Education, Number 118, Summer 2002*, Jossey Bass
- Blake, N., Smith, R. & Standish, P. (2013), 'The universities we need: higher education after Dearing', *Routledge*
- Bloom, B.S. (1981). 'All Our Children Learning', *New York: McGraw-Hill*
- CHERPA Network (2011), 'The design & testing the feasibility of a multi-dimensional global university ranking', *U-Multirank conference, Brussels, Thursday 9 June* (<http://www.ireg-observatory.org/pdf/U-MultirankInterimReport.pdf>)
- Chickering, A.W. & Gamson, Z.F. (1987), 'Seven Principles of Good practice in Undergraduate Education', *AAHE Bulletin, March*
- Ewell, P.T. (2013), 'The Lumina Degree Qualifications Profile (DQP): Implications for assessment', *Occasional Paper #16, NILOA*
- Gott, T. (1999), 'Best practice benchmarking', *Lecture notes, Academic Audit Unit, Wellington*
- Guthrie, H. (2009), 'Competence and competency based training: What the literature says', *National Centre for Vocational Education Research (NCVER), Adelaide*
- Hall, C. & Woodhouse, D. (1999), 'Accreditation and approval in New Zealand: Major surgery for the national qualifications framework?', in *Fourie, M., Strydom, A.H. & Stetar, J. (eds) Reconsidering quality assurance in Higher Education: Perspectives on programme assessment and accreditation, The University of the Orange Free State, Bloemfontein,*
- Hussey, T. & Smith, P. (2002), 'The Trouble with Learning Outcomes', *Active Learning in Higher Education, November, Vol 3, 3, pp220-233*
- James, R., McInnis, C. & Devlin, M. (2002), 'Options for a National Process to Articulate and Monitor Academic Standards Across Australian Universities', *Submission to HE Review, Department of Education, Science and Training, Canberra*
- Marginson, S. (2006), 'Australian universities in the global context', *Campus Review, 22 March, pp8-9.*

Quality Assurance Agency (1996), 'Graduate Standards Programme: Final Report', QAA

Rhodes, T.L. (2010), 'Assessing outcomes and improving achievement: Tips and tools for using rubrics', AAC&U

Sadler, D.R. (1987), 'Specifying and Promulgating Academic Achievement Standards', *Oxford Review of Education*, Vol. 13, No. 2, pp191-209

Scott, I. (2009), 'The Learning Outcome in Higher Education: Time to think again?', *Worcester Journal of Learning and Teaching*, Issue 5

Stella, A. & Woodhouse, D. (2006), 'Ranking of higher education institutions', AUQA Occasional Publication No. 6, August 2006, Australian Universities Quality Agency, Melbourne

Usher, A. & Savino, M. (2006), 'A world of difference: A global survey of university league tables', Educational Policy Institute, *Canadian Education Report Series*, Toronto

Woodhouse, D. (2000), 'Models for quality improvement', Presentation to INACAP's VIth Annual Seminar on Educational Innovation: Quality Assurance in Education, Santiago Chile, 24 August