

Can we create a more strategic approach to performance indicators and standards in Australian higher education?

Presentation for the University of Melbourne policy seminar series
Investing in the Future: Renewing Tertiary Education
4 August 2008

Richard James
Centre for the Study of Higher Education
University of Melbourne

Introduction

First let me thank my colleagues in the Centre for the Study of Higher Education and the LHMartin Institute for Higher Education Leadership and Management who have been involved in planning this series of seminars: Leanne Howard, Michelle Segal, Russell Smith, Karla Fallon and Professors Simon Marginson and Lynn Meek. We're honoured to have this opportunity to engage with and to contribute to the discussions around the Bradley and Cutler Reviews and exceptionally pleased with the press coverage the series has received in Campus Review, The Australian and The Age in recent weeks.

Standards and performance measurement — in particular the quantification of performance — are now *critical* issues for the tertiary sector. We must set-up Australian higher education for a more competitive international market in higher education — a market hungry for information — and ensure the international standing and portability of Australian awards. For this we need fresh approaches to performance measurement and a more sophisticated and stringent approach to standards.

This evening I will outline some of my ideas for the collection and use of data in higher education, especially with the prospect of mission-based compacts before us. I will propose broadbrush directions in which we need to progress as a sector, without presenting a detailed model. In developing my points I will focus in the main part on teaching and learning.

It is so complex to discuss standards and performance indicators in the current HE context. And, it sounds like such a dry and boring topic: last week when Simon foreshadowed this evening's seminar he said it would be 'more technical', surely the kiss of death for any speaker. So I'll try to make this presentation as entertaining as possible and it certainly won't be overly technical.

The elusive nature of higher education standards: It's not the Olympic games

Before developing my ideas, let me stress four points that are probably well known to all of you:

- the act of measurement does not in itself lead to improvement, but it does much to signify what is valued. This can have significant effects on behaviours ... performance indicators have to be used very wisely;

- many of the things we value most in higher education are exceedingly difficult to measure and to quantify, thus we tend to resort to proxy measures. There is the ever-present risk of reducing a complex human enterprise to superficial indicators — we have to resist the notion that everything can be quantified;
- the concept of standards is highly elusive in higher education and the market in itself cannot guarantee standards — some regulation is needed.

Oddly, standards are one of those things that we struggle to define with precision yet we tend to know when they're there and when they're not. Despite the everyday use of the term 'standards' we are some way from a technical definition. As a broad principle, much of what is referred to loosely under the rubric of 'standards' and 'protecting standards' relates to guarding inputs and processes rather than understanding the quality of outcomes.

Now, as it turns out it's quite timely to be discussing performance and standards with the opening of the twenty-ninth modern Olympics only four days away. With most – but not all - of the sports in the summer Olympics the measurement of performance and standards is objective, transparent and immediate:

- excellence leaps out.
- the rules and goals are absolutely clear. Everyone starts from the same point and the starter's gun is especially sensitive to diversity: it works across all contexts and speaks all languages.
- the metrics are well understood ...
- and fine grained differences in performance can be measured.
- plus, in the main part athletic performance is beyond manipulation and opaque behind the scenes deals — though not entirely so, as we're reminded of a little too regularly.

So, at the Olympics at least, performance can be measured fairly unambiguously. What constitutes 'world-class' and 'excellence' is generally quite clear-cut.

Let's compare this with higher education. Obviously the outcomes are far less tangible and far less immediate. Further, the goals and desired outcomes are contested. Excellence is less absolute, it is located within contexts and shifts over time. At least to some degree, quality in higher education is subjective and lies in the eye of the beholder.

It could be argued that in a mass or universal higher education system if 'excellence' is to have meaning then it involves achieving a good 'fit' between the needs of particular groups of students and the education that is provided.

My point here is simple: the measurement and comparison of institutional performance in higher education is imperfect and somewhat fraught. And thus the use of PIs for improving performance, or for rewarding or providing an incentive for enhanced performance, is an imperfect science.

The elusiveness of standards has some interesting effects. First, it allows us to make, and to get away with, exaggerated claims. Every university can be a leader, for this claim can seldom be decisively rebutted (or defended). We have become extremely successful in 'talking up' the quality of Australian higher education. Expressions such as 'world-class', 'internationally recognised' and 'international

standing' are repeated mantra-like on university websites and in promotional literature.

Second, almost conversely, it allows academic standards to be the site of cheap polemic. We're all familiar with the headlines that appear on an almost weekly basis around the world: 'standards in decline', 'standards being eroded'. These claims are equally difficult to defend or rebut, as the case may be. Our notions of standards, despite being based on powerful academic intuitions and experience, are slippery and elusive – it is often difficult to point to objective evidence one way or the other. 'Standards' is an everyday term but we are some distance from a technical understanding of precisely what standards might be.

But focus on standards we must

Acknowledging this complexity does not mean we shouldn't strive to measure performance in higher education or that we shouldn't monitor standards. Of course we must be concerned with both, hence my argument that we need to develop more sophisticated performance measures and a more explicit framework for understanding the nature of standards and monitoring them.

Let me highlight three salient aspects of the present context that justify this position:

- first, the continuous improvement of teaching, learning and educational outcomes will depend greatly on our capacity to develop evidence-based approaches to planning and resource allocation — the education industry, across all sectors, is going through the somewhat painful process of learning how to take evidence seriously and to implement evidence-based approaches.
- second, the OECD feasibility study *Assessing Higher Education Learning Outcomes* (AHELO) has the potential to have profound effects internationally, especially on the metrics for university rankings (which are presently based largely on research) with resultant effects on international student flows. If AHELO eventually gets underway — and I imagine it will in some form, for so many stakeholders want it — it could unsettle overnight the faith we have in the quality of Australian degrees.

The feasibility study has four strands:

1. The assessment of generic skills
2. The assessment of discipline-specific skills
3. Measurement of the value-added or contribution of TEs to students' outcomes
4. Contextual measures and indirect indicators of tertiary education quality.

The second and third of these will break quite new ground and are particularly important for Australian higher education.

- third, and far from least, the success of mission-based compacts — should we head in this direction — will rely heavily on the government's and the sector's confidence in the validity of the measurement of each institution's performance against its negotiated goals — in fact, performance measurement is a central element in a compacts approach, as Michael

Gallagher stressed in a previous seminar. I will return to compacts in a moment.

How good are the standards of Australian higher education awards?

Just how good are Australia's academic standards? The simple answer is that it is difficult to know for sure. I don't think we should be too worried for there are broadbrush indications that academic standards should be high: we have a strong national QA framework with the national protocols, AUQA, extensive data collection, such as the deployment of the CEQ, and so on. Universities have solid governance and management arrangements, academic staff are well-qualified and the Australian Learning and Teaching Council (ALTC) and its predecessors have supported the enhancement of teaching and learning.

But there are gaps or signs of 'fragility', as Richard Larkins described them:

- universities are starved for resources;
- less than optimum staff/student ratios and the casualisation of undergraduate teaching;
- a trend towards 'fast tracking' completions;
- pedagogical pressures created by the number of international students;
- an apparent degree of student disengagement;
- problems of student income support and the patterns of excessive paid work; and
- a well-performing but uneven school system, which of course underpins what our universities can achieve.

Vice-Chancellors have had the unenviable task of arguing that universities are under-resourced yet at the same time claiming standards have not been compromised. This kind of juggling can't go on much longer.

My guess is that Australian undergraduate programs will stand up well to international standards and some of our honours program are probably of the highest international standards — honours in some ways is the high watermark in Australian higher education. I believe Masters programs are more problematic and are probably highly uneven in quality. Australian PhD dissertations probably also vary somewhat in quality but where there is intensive international examination the standards are undoubtedly sound. However, much hunch is that our PhDs graduates are not as broadly educated as the best of US graduates and we might need to rethink the structure of the PhD in Australia in order to develop a wider range of skills — and of course some universities are already doing this.

But all of this is mere guesswork. The devices we have for setting and monitoring standards, especially across institutions, have limitations. The AQF offers only a modest description of what constitutes a bachelors degree and has failed to serve as a significant reference point within universities. Further, we have operated on a highly devolved assessment regime in which, by and large, it is assumed that the sum of the parts will generate the whole. We do not have a tradition of 'capstone' testing, nor of cross-institutional moderation of external examination, except for research higher degrees of course. Even the grading and reporting of student learning is not standardised in Australia, with no agreed system-wide GPA and individual universities having their own sometimes arcane grading nomenclature. The ACER's Graduate Skills Assessment (GSA) is the closest we've come to testing

that would allow comparison across universities, however the GSA has not gained traction within the sector.

In its second cycle of audits AUQA is examining academic standards at the request of the previous Minister. AUQA has a challenging task for certain pre-conditions do not exist for understanding the character of standards on an institution-by-institution basis without some overlay of comparative data. Sufficient data do not yet exist.

How well do we use performance data?

The Australian higher education system has an impressive record — indeed an innovative record — in systematic, system-wide data collection and reporting dating back at least to the research of Russell Linke over 20 years ago. In the main we have used performance data with a ‘light touch’ and data have not been used for performance funding, but for the usual ‘volume’ and ‘throughput’ measures.

The Learning and Teaching Performance Fund has been the most notable exercise in performance-based incentives. The LTPF has been a controversial exercise. The LTPF has been welcomed by some senior academics for it has assisted them in their quest to draw attention to teaching issues and to marshal energies around enhancement, and these are good things. The additional discretionary resources, for the institutions that have received them, have been welcome too.

But the LTPF is deeply flawed if we examine its value as a tool for continuous improvement. First, the data are extensively lagged. Second, the processes used to devise the rankings and to award the funds have served to exaggerate small differences, differences that may not count for much in practical terms. Third, and primarily, the indicators are dubious measures of institutional performance. Graduate employment rates are entwined with institutional reputations and, in some cases, course reputations, and thus are not an adequate measure of student learning. Worse, the inclusion of the Generic Skills Scale from the CEQ means we use graduate *self-reports* of their generic skill acquisition as an indicator for the allocation of LTPF funds. It is ironic that we seldom consider student self-assessment for grading purposes but find it to be acceptable for the competitive allocation of sizeable sums of public money.

With the LTPF the indicators were chosen in large part because these are the only cross-institutional data we have not because they are believed to be good measures of what the LTPF claims to recognise. Now there is an argument to the effect that we shouldn’t be too concerned about these flaws, that the LTPF has kept a spotlight on teaching and that this in itself is sufficient. I don’t accept this — I believe this is shallow thinking. We need to get to move to a point where evidence-based decision-making is taken seriously in higher education. For this we need valid and reliable indices.

Looking to the future: Achieving progress on standards, performance measurement and the use of indicators in a mission-based compacts environment

I believe we now face significant challenges in the areas of standards and performance indicators if these are to match our aspirations for the quality of the sector. Compacts will help set the pre-conditions for a more diverse and responsive

higher education sector. For this to be successful, however, our notions of standards and performance will need to adjust accordingly. We need to choose and use performance indicators very carefully. We need to develop a performance indicator framework that supports and does not undermine differentiation and which does not nurture the present cycle of vertical stratification based on reputation and positional status.

In describing the idea of compacts in March of this year, Senator Kim Carr said:

Universities will have a reciprocal responsibility to explain their purposes, and to report publicly on how well they have performed against their own goals and expected performance standards’.

What might this look like in practice? Here there is much devil in the detail but we can assume with some confidence that the government will not be comfortable handing over the entire measurement of performance to the institutions themselves. There are a number of questions

The unit of analysis?

How will we judge the performance of the sector as a whole versus the performance of its component institutions? How will institutional performance be compared? Indeed, is there any need to compare institutional performance in a compacts model?

The indicators?

Will there be core, common performance indicators? Is it possible to establish an indicator framework that does not have a gravitational pull towards bland institutional uniformity?

Use of the data?

Will competitive, performance incentive funding, such as the LTPF, have any value in a mission-based compacts environment?

I believe for compacts to be successful and for the sector to have confidence in a compacts-based funding model we need to:

- develop agreed sector-wide metrics, based on common definitions, for measuring performance in the three main areas of performance: teaching & learning, research & research training and community engagement or knowledge transfer.
- develop metrics that shift the emphasis to *outcomes/impact* measures of performance to complement input/process measures – a two-tiered approach to performance indicators is necessary in which, where possible, the focus is placed squarely on outcomes measures. This is the challenging conceptual breakthrough we have to make.
- establish core performance measures for all institutions, regardless of negotiated mission — for some outcomes will remain common to all — and additional indicators to be used as appropriate on the basis of institutional mission (e.g. regional contribution);

- Develop a *value-added* conception of institutional performance in the area of student learning outcomes and adopt both *absolute* and *value-added* measures as legitimate indicators of outcomes.

More broadly, the HE sector needs to explore the possibilities for

- developing a better articulation of the *minimum* academic standards for Australian degrees, as opposed to aspirational standards. The target should be that Australian degree programs produce educational outcomes comparable with the top-end OECD countries. Strengthening the Australian Qualification Framework (AQF) would be part of this process — the AQF must be enhanced if it to be of value to the higher education sector . Professor Jack Keating has recently led a project team to develop a new AQF that has recently reported to DEEWR.
- strengthening cross-institutional benchmarking within kindred fields of study.
- developing more international reference points or benchmarks for Australian academic standards.

So, in summary, I am proposing a performance indicator framework in which:

- there is an agreed national indicator framework;
- standard system-wide definitions are developed for all performance indicators;
- outcomes indicators are distinguished from process indicators;
- A priority is attached to outcomes indicators;
- core PIs are identified that apply to all institutions regardless of negotiated compact; and
- additional, mission-specific PIs are identified that are appropriate for certain institutions on the basis of their distinctive missions and goals.

Within this model it might be quite reasonable in some areas to put universities into head-to-head comparison: some desirable outcomes ought be spread roughly evenly across the sector, for others, however, we might expect institutions to specialise.

What might this indicator framework look like in practice?

The following table is merely an illustrative sketch ... obviously it is not fully populated with data elements, though I have indicated what might be five core outcomes indicators for teaching and learning.

A first sketch ...

Two tiers of Pls

	Outcomes/Impact 'Performance indicators'	Inputs/Processes 'Markers'
Teaching and Learning	Graduate skills Completions Employability Equity	AUSSE CEQ etc.
Research and Research Training	(wait for ERA!)	Research income etc. etc.
KT, Community engagement — '3rd stream'	???	Range and volume of activities etc. etc.



A first sketch ...

	Outcomes/Impact 'Performance indicators'	Inputs/Processes 'Markers'
Teaching and Learning	Graduate skills Completions Employability Equity	AUSSE CEQ etc.
Research and Research Training	(wait for ERA!)	Research income etc. etc.
KT, Community engagement — '3rd stream'	<ol style="list-style-type: none"> 1. <u>Undergraduate</u> knowledge and skills Absolute and Value-added Generic and Discipline-Specific 2. <u>Graduate</u> knowledge and skills 3. Completions 4. Employability 5. Equity: Low SES, Indigenous people 	Range and volume of activities etc. etc.



Much technical work would be needed to achieve a framework of this kind. The assessment of teaching and learning will present even more challenges than the assessment of research — and we know from RQF/ERA how difficult this can be and how much anxiety is created.

As for knowledge transfer, community engagement or ‘third stream’ activities, it’s fair to say we do not, as yet, have the metrics to do this at all well. There is not a long tradition of measurement in this area. Of course it would be possible to use audits and ‘volume of activity’ measures but these are imperfect proxies for outcomes and the contribution

There is no point in developing this framework any further this evening, for we will all have different views on its components. Mike Gallagher proposed the formation of a transition advisory group to work through new policy and financing arrangements. If such a body were formed one of its first objectives would be to articulate the way in which performance measurement is to take place within the sector. It is highly important that such performance data be objective and transparent. The indicators need to be agreed to and resistant to manipulation. Importantly, I see no reason why different metrics would be needed for individual institutions. The essence of the compacts idea lies in the differing mix of activities rather than in particular activities being unique to particular institutions.

A case study: Measuring performance in teaching and learning

To illustrate my argument let me focus in more detail on teaching and learning. In teaching and learning much of our attention at national level has focussed on input and process measures rather than outcomes measures. But outcomes will become more prominent. We can confidently predict that:

- Curricula will diversify, including through sandwiching of work experience and study.
- People will dip in and dip out of higher education and the demand for recognition of prior learning will grow
- Students’ patterns of engagement (and disengagement) with university will grow more complex.

What makes higher education ‘higher’ will become even less certain. In a context of diversifying curricula and modes of delivery and student engagement it is what students have learned that counts not how they have learned it. The assessment of student learning — that is, the outcomes or impact of university education — will become the most reliable bedrock for determining the effectiveness of the sector and of individual institutions. This is where AHELO might become so significant.

The present indicators of teaching and learning are imperfect proxies for what is of most interest to us. The table to follow is an attempt to illustrate this in simple terms.

The CEQ has been a useful management tool for focussing attention on the quality of teaching. But the CEQ is flawed as a performance indicator for comparing learning between institutions and programs because it reveals nothing about the depth or breadth of student learning. The CEQ could be deployed with Year 12 graduates and university graduates and the results would be barely discernible if the groups were receiving teaching they perceived to be of the same quality. We have tended to

use the CEQ as a proxy for learning outcomes, which clearly it is not — the CEQ tells us little about student learning in absolute terms or in value-added terms.

The contrasting approaches:

Course Experience Questionnaire (CEQ)	Graduate perceptions of teacher behaviours and their own beliefs about their learning	A process measure ... a highly indirect measure of learning. <u>Evidence for the claimed associations is weak</u>
The ACER's AUSSE	Students' self-reports of their study-related activities	A process measure, but an important one ... an indirect measure of learning. <u>Stronger evidence for association with learning outcomes</u>
Assessment of student learning	The direct assessment of what students/graduates know and can do.	A true outcomes measure ... a direct measure of learning. 'Science' of assessment in HE not well developed.

The Australian Survey of Student Engagement (AUSSE), developed by Hamish Coates of the ACER, with its focus on purposeful student activities related to learning moves us closer to a measure of student learning outcomes but still does not directly measure this. Nonetheless, the AUSSE has the potential to be a superior, more timely indicator of future outcomes. The challenge here, however, is convincing academic communities to embrace the AUSSE as a performance measure when it can be argued they have limited influence over what students are doing.

The complications:

Course Experience Questionnaire (CEQ)	Graduate perceptions of teacher behaviours	Relatively easy to implement but data are heavily lagged.
The ACER's AUSSE	Students' self-reports of their activities	Relatively easy to implement, data far less lagged.
Assessment of student learning	The direct assessment of what students/graduates know and can do.	Challenges in implementation, 'ownership' of responsibility

What if we measured the performance of Olympians in this way? The table to follow offers some suggestions. This comparison of our present situation with the Olympics is flippant of course but it does make a serious point.

Compared with the Olympics?:

Course Experience Questionnaire (CEQ)	Graduate perceptions of teacher behaviours	Athletes are surveyed on their coaches' skills and the training 'climate' they've created - no need to run the races.
The ACER's AUSSE	Students' self-reports of their activities	Athletes report their training regimes and training data, from which performance is inferred - no need to run the races.
Assessment of student learning	The direct assessment and reporting of what students/graduates know and can do.	The races are run.

Final comments

Ironically, as we move closer to the indicators *that really count* the more the 'co-produced' character of higher education becomes highly apparent and 'ownership' of responsibility becomes problematic, as the following table indicates. The academic community has been somewhat resistant to some of the measures used to monitor performance and depicted these in terms of managerialism and the like. These tensions are unlikely to be any less into the future.

Outcome/impact measures	Process measures
Relationship to day-to-day academic teaching not always clear	Can be directly related to day-to-day academic activities
May reflect the co-produced character of education outcomes, i.e. the result of a mix of staff and student efforts,	Can focus squarely on the contribution of academic staff to the co-production of outcomes
... if used as performance indicators likely to meet with more resistance and less 'ownership' on the part of academic staff	... when used as performance indicators have been quite useful for management purposes

There is an inexorable trend towards more standardised, more independent testing in higher education. Few would argue that such testing can tell us everything we wish to know, but it can tell us something. Standardised testing in the school sector has been the source of much criticism from within the teaching profession, of course, where the collection of data has been seen variously to threaten professional judgement, to narrow what is valued in learning and to homogenous curricula. Whatever approaches are taken in higher education we must be sure not to 'deprofessionalise' academic work. Whatever directions are taken at this point we need to be mindful of how we bring our academic communities with us. Academic leadership will have a prominent role.