1. Introduction

Universities around the world are regularly assessed in terms of their performance and their relative positioning on the world stage. In many countries, a significant proportion of university revenue is derived from government funding sources, so it is not surprising that funds are generally allocated on the basis of perceived performance. This is particularly true in the case of research funding, which is often competitive in nature.

The challenge for governments, benefactors and universities is to determine what constitutes good performance in a university; how this can be systematically compared to other universities (often across multiple countries) and whether it is even possible to measure performance meaningfully at all.

Universities tend to deal in abstract concepts and knowledge that may not deliver tangible outcomes for years, decades or even centuries. However, they are large annual consumers of funds which need to be regularly justified and reviewed by governments and other benefactors. Additionally, students who are paying tuition fees have a genuine interest in knowing exactly what they are getting for their money, beyond the abstract concept of "an education".

Universities generally collect large amounts of data about their operations and because there are common elements to the data which is collected, these have often emerged as metrics by which universities the world over are assessed.

The two broad sets of metrics which are employed to assess universities relate to research and to teaching. Additionally, some organisations also endeavour to evaluate the “esteem” in which institutions are held, notwithstanding the fact that these sorts of measures tend to be arbitrary and highly subjective in nature.

Needless to say, because not all institutions will perform well across all metrics, the specific metrics that are used to assess universities will directly impact on the way in which institutions appear to perform. A university which performs well across one set of metrics may perform more poorly against another set.

The large number of performance metrics that are available to assess university performance lead various groups to rank universities either nationally or internationally. Each ranking organisation has its own unique set of metrics and so the results for a particular institution can vary markedly from one set of rankings to another.
University rankings tend to be skewed towards either research metrics or teaching metrics and so it is quite possible that a university which ranks poorly on research metrics may still rank quite well on schemes that focus on teaching.

Importantly, it needs to be remembered that universities are fundamentally places of learning – the staff are there to learn and the students are there to learn – they are not principally founded as places of teaching, even though it is inevitable that universities have to undertake some teaching roles in basic principles – for example, in mathematics or physics. With this in mind, it is not surprising that many university rankings focus on research metrics (i.e., the ability of the institution to learn and create knowledge for others to learn) and less are focused upon teaching metrics (i.e., those focused on the transfer of accepted knowledge). To complicate matters further, teaching metrics tend to be more abstract and inconsistent than research metrics so there is naturally a greater emphasis on those parameters which are more consistent and defensible in a ranking scheme.

2. Limitations of Metrics

University metrics are no different to key performance indicators used in business and industry in the sense that they have a self-fulfilling outcome which is not always desirable. When an individual or an organisation is assessed based upon metrics they will naturally respond by trying to optimise those metrics. In a university, if an individual is judged on the number of research publications or citations, then he/she will seek to maximise them. All measurement systems end up distorting the system which is being measured because they change the intrinsic behaviour of that system.

In basic terms, there isn’t a problem in an individual or organisation responding to metrics by changing their performance – for that is the intention of the metrics. The issue, however, is that in some organisations the role of individuals is far more complex than that encapsulated by the metrics. For example, in a university, a professor may be judged on publications and citations but his/her role needs to be much broader than that if the university is to function. A good professor might need to mentor and counsel students; undertake administrative duties; participate in extra-curricula activities; support university committees, and so on. By applying metrics as the key criterion for performance assessment, all the other activities are automatically relegated to secondary importance. In the extreme case, an organisation that simplistically applies them can become completely dysfunctional because anything not covered by metrics is either ignored or handled poorly.

In business and industry, metrics are effective performance tools because they can be linked directly to tangible outcomes and productivity. For example, if someone is required to produce 100 screw-drivers a day, then it is easy to determine whether the screw-drivers have been produced, and it is easy to have subsequent metrics that ensure that the screw-drivers meet quality standards. Unfortunately, this approach doesn’t translate well to the university system.

A good academic may be able to write one research paper in a year. However, if the performance metric is to maximise the number of papers, then that academic may choose to achieve his/her ends by writing five low-quality papers, or even by dividing his/her single paper into multiple parts and publishing those separately. How can this be avoided? In universities another set of metrics are applied – citations. If a paper is cited many times then it is assumed to be a quality paper. The application of this secondary metric doesn’t actually resolve the problem because it just creates another point of distortion – an academic may choose to ask colleagues or students to cite his/her papers – or the same academic may sit on the editorial board of a journal and insist that authors cite his/her papers as a precursor to acceptance for publication.

The key point here is that university outputs are abstract and difficult to quantify with metrics and, when metrics are applied, they inevitably lead to unwanted distortions and misfeasance which becomes increasingly difficult to identify as each subsequent metric is deployed to overcome the distortions of the previous one.
The most significant shortcoming of metrics in assessing universities is encapsulated by the Latin phrase “post hoc ergo propter hoc” (which literally means “after this therefore because of this”). A university which performs well across all its metrics may well be highly regarded in the international arena. But the key question is whether the university is highly regarded because it performs well in all its metrics?

In the university system, it is often the case that a great university performs well in all its metrics because it is a great university – and the key (and often intangible) reason that a university is great does not relate to the metrics at all but, rather, to the transformational changes that it has made to science and society.

The assumption that metrics do indeed relate to the transformational change in science or society is not always a valid assumption in a university. Indeed, it is possible for a university to perform well across a wide range of metrics without ever having made a transformational change – in those cases, the metrics are not an accurate reflection of the system they are endeavouring to model.

3. Research Metrics

At a basic level, research is inherently easier to measure and model than teaching simply because there are so many metrics available for use. As with all metrics, the fact that something can be measured does not necessarily mean that it has any intrinsic meaning. The challenge in research metrics is to understand the purpose and significance of each metric and the distortions that arise by applying it as a performance measure for individuals or institutions.

Table 3.1 summarises the basic metrics that are employed and their limitations.
<table>
<thead>
<tr>
<th>Research Metric</th>
<th>Objectives/Rationale</th>
<th>Systemic Distortions Created by Application of the Metric</th>
</tr>
</thead>
</table>
| Publications           | Research publications are generally subjected to peer review, often at an international level. The more publications that are accepted in peer reviewed journals the more likely it is that the academic is delivering an ongoing stream of internationally accepted research outcomes. | • Large number of peer reviewed journals appearing with decreasing standards  
• Increasing numbers of frivolous publications that satisfy decreasing journal standards  
• Staff bullying research students and junior researchers to publish more  
• Research publications become little more than peer reviewed technical reports rather than documents detailing significant research breakthroughs |
| Citations              | Publication volumes are insufficient to assess the impact of research. The number of times a paper is cited by other researchers gives an indication of the broad acceptance of the research.                                                                                       | • Research can be highly cited because it is bad as well as because it is good  
• Academics can ask colleagues and students to cite work to improve their metrics  
• Academics acting as journal referees or editors can insist upon authors citing their work to improve their metrics  
• High capital cost research and large epidemiological studies are cited highly because it is difficult to reproduce rather than because they are transformational |
| Awards                 | Awards reflect the highest standards as determined by reputable peer review panels and are on a competitive basis. Awards can recognise transformational research.                                                                                                       | • Major research awards require significant lobbying from home institutions because of the prestige they bring to the institution  
• Major research institutions tend to be overrepresented on award panels so the awards tend to be self-fulfilling |
| Competitive Research Grants | Competitive research grants are awarded based upon peer review of the history and strength of the academics who receive them.                                                                                       | • Tend to create cliques in grant review panels  
• In order to give the appearance of fairness, granting processes have to take a formulaic approach and therefore subdue transformational and high-risk research in favour of incremental research. |
| Research Student Load  | Research students are the engine room of university research – the larger the engine room the greater the research capacity.                                                                                                         | • Student load can be increased by recruiting low-calibre students |
| Research Student Completions | Research student completions are an indicator that peer reviewed learning outcomes are taking place.                                                                                                               | • Academics can have colleagues examine research degrees |
| Eminent Alumni         | Alumni who move on to achieve greatness reflect well on the institution’s research.                                                                                                                                      | • Assumes that alumni eminence occurs because of the source institution rather than despite it |

Table 3.1 – Research Metrics and Systemic Distortions
4. Teaching Quality Metrics

Teaching quality is an abstract concept and it is not always reasonable to measure teaching quality at the end of a course or degree program – sometimes it may take years or decades for students/graduates to appreciate the significance of the teaching at universities. It is also important to keep in mind that universities are places of learning rather than teaching. A student or graduate who has learned despite the quality of teaching may have achieved more from his/her university education than one who has learned because of it.

Table 4.1 shows the various metrics that are commonly employed and the distortions they bring to the system. Overall the level of distortion in these metrics is so great that it is difficult to use them with significant confidence.

<table>
<thead>
<tr>
<th>Teaching Metric</th>
<th>Objectives/Rationale</th>
<th>Systemic Distortions Created by Application of the Metric</th>
</tr>
</thead>
</table>
| Completion Rates| The higher the proportion of commencing students that complete a course or degree the better the teaching | • High completion rates may be a function of “dumbing down” of courses/degrees  
• A low completion rate may signify that teaching and assessment standards are more rigorous |
| Student Satisfaction Surveys| Students/graduates are the best judges of the quality of teaching | • Students/graduates may not understand or appreciate the quality of what has been taught for years or decades  
• Students may score easier subjects more favourably than challenging subjects – leading to “dumbing down” of courses |
| Employability Statistics and Employer Perceptions| The more employable that graduates are the better the university teaching must be | • Universities which attract high calibre students and have poor teaching may still have better employability outcomes than universities which have low calibre students and good teaching  
• Employers often focus on readily usable skills so applied universities may have better employability outcomes than those who create thinkers that are better long-term employees |
| Student Entry Scores| The higher the entry scores for an institution the better its teaching must be | • Entry scores may be completely unrelated to teaching quality and more related to perceived institutional prestige or availability of particular courses |
| Student Demand| The higher the demand for an institution the better its teaching must be | • Student demand may be completely unrelated to teaching quality and more related to perceived institutional prestige or availability of particular courses |

Table 4.1 – Teaching Metrics and Systemic Distortions
5. University Rankings Based Upon Metrics

<table>
<thead>
<tr>
<th>Ranking</th>
<th>Key Metrics</th>
<th>Comments</th>
</tr>
</thead>
</table>
| Shanghai Jiao Tong University Academic Ranking of World Universities (ARWU) | • Quality of Education (Alumni)  
• Quality of Faculty (Nobel Prizes, Fields Medals, Highly Cited Authors)  
• Research Outputs (Publications)  
• Per Capita Performance | • One of the most respected international rankings which ranks the top 500 universities from a selected pool of around 1,000.  
• The Alumni and Major Award criteria prevent universities from artificially working their way up the rankings simply by increasing the volume of research outputs – they have to have a history of transformational achievement  
• The ARWU is somewhat skewed towards scientific institutions so American universities feature prominently and European universities less so |
| Times Higher Education (THE) World University Rankings | • Teaching  
• Citations  
• Research (Volume/Income)  
• International Mix (Students/Volume)  
• Industry Income | • THE rankings methodology has been updated for 2010 and is significantly different to previous ranking approaches  
• Each of the five metrics categories has 13 indicators  
• Previous rankings had heavy weighting on opinion polls and were regularly criticised for this – the 2010+ rankings have a lower weighting to polled perceptions  
• The new formula allows universities to artificially improve their performance by increasing performance on metrics without transformational change – the Shanghai Jiao Tong has retrospective safeguards such as alumni, etc. |
| Leiden University Rankings | • Publications and Citations  
• Normalised Citations (Citations divided by Impact Factor) | • Straightforward bibliometric ranking system  
• Based on publications and citations  
• Uses normalised citations which adjust for field impact factor  
• Limited to ranking of university research performance  
• Assume that citations and transformational changes to science and society are correlated |
| QS Top University Rankings – Quacquarelli-Symonds | • Academic Peer Review  
• Employer Review  
• Citations Per Faculty  
• Student Faculty  
• International Faculty  
• International Students | • QS was the former supplier of data to the THE rankings above  
• QS rankings contain a number of perception based metrics derived from surveys – these appear to be more favourable to English-speaking and European universities than Asian universities. |

*Table 5.1 – Major International University Rankings and Limitations*
6. Research Quality Assessment Exercises

The limitations and distortions associated with simplistically applying metrics (particularly bibliometrics) to the assessment of research have led various governments to attempt to have a more mature assessment process for research excellence. These attempts include the British Research Assessment Exercise (RAE) and the Australian Excellence in Research Australia (ERA) exercise. The basic concept is that universities submit their best research work across a number of fields and these are intelligently assessed by field panels. The potential benefit of such an approach is that genuine excellence can be differentiated from blatant attempts by individuals and institutions to manipulate metrics. Essentially the assessment exercises provide a peer review of research in each field for each university.

There are numerous problems with these exercises. Firstly, it takes a large amount of time and effort for institutions to collect and assemble the sort of data that is required for such an exercise. Secondly, there is the issue of assembling expert panels to assess the work. Thirdly, once each university has submitted a portfolio of research work for assessment in each field, there is an issue of how to establish a consistent and fair approach to assessing the research. Ultimately, because it is impractical to get agreements on a fair approach across many institutions with competing interests, the result tends to be that panels do little more than aggregate bibliometric data and make similar conclusions to those that can be derived in an automated fashion. Fourthly, most countries have competitive research grant processes which already provide a vehicle for the sorts of peer review that are provided in research assessment exercises.

Finally, the most profound shortcoming of research assessment exercises is the amount of time required to conduct them. It takes universities a year to prepare and collect data, and six to twelve months to convene expert panels to analyse the data. By the time the results are returned to universities, some two years have elapsed. The publication date of the most recent publications that have been reviewed is then approximately three years old. Given that it takes one to two years to publish research (from time of submission), the research assessment exercise ultimately ends up reporting on research that was undertaken by universities four to five years earlier. With an academic staff turnover of, say, 10% pa, a research assessment exercise effectively compiles a report on institutions that no longer exist in the form they were when the data was collected.

7. Structural Rankings

It is possible to create basic rankings of universities which have some integrity, longevity and resistance to manipulation by individuals and institutions. The simplest way to do this is to increase the scale of the metrics that are deployed to eliminate micro-manipulation. It is much more difficult to artificially manipulate macro scale parameters than micro parameters.

There are three basic structural parameters which reflect the overall performance of a university. These are:

(i) **Total number of academic/research staff**

Regardless of which micro metrics are used (e.g., publications, citations, etc.), it becomes evident that over a large number of staff there is largely a linear relationship between performance and staff numbers. Of course there are always some staff who publish more and who receive more citations but – overall – both nationally and internationally the relationship is linear. Additionally, a large number of staff add to the intellectual richness and diversity of a university. Most importantly a large number of staff create critical mass which can be used to minimise academic misfeasance. For example, a physics department with 100 academic staff is more likely to be able to identify academic errors or inconsistencies in a staff member's work than a department with only five staff.
(ii) **Total non-recurrent assets of an institution**

In order to provide an environment for quality research and learning an institution needs to have assets including buildings, laboratories, endowments for scholarships, etc. In scientific disciplines these assets need to be infrastructure based - in areas such as business the assets may be staff who need to be paid significantly more because of the external professional environment. It is generally the case that there is a linear relationship between university assets overall performance in learning and research

(iii) **Student:staff ratio**

From the perspective of a student, the most important element of a learning environment is to have access to a depth and breadth of expertise from which one can draw knowledge. A university which has a low student:staff ratio can offer both depth and breadth of knowledge to students. Additionally, because there are fewer students per staff member, staff have more time available to create new knowledge through research. There is generally an inverse relationship between overall institutional performance and student:staff ratio. The world’s leading universities maintain low student:staff ratios.

At both a national and international level, universities which have a high number of academic/research staff, a large non-recurrent asset base and a low student:staff ratio perform better than those that don’t. In some institutions (indeed some countries), the student:staff ratio reporting is artificially improved by including sessional (part-time) staff who do not add directly to the knowledge depth/breadth of the institution.

Notwithstanding the anomalies that can arise in student:staff ratios, the three structural parameters are difficult to artificially manipulate because they are large in scale and readily validated – and, if they are manipulated, it is a relatively straightforward task to correct them.