

Twenty Years Of University Report Cards: Where Are We Now?

(Version submitted to *Higher Education in Europe*; final article forthcoming: Vol. 30(2)
June, 2005)

Dr Nina Van Dyke
Level 7, 161 Barry Street
The Melbourne Institute of Applied Economic and Social Research
The University of Melbourne
Victoria 3010
Australia
T: +61 3 8344 2116
F: +61 3 8344 2111
E: n.vandyke@unimelb.edu.au

Abstract

It has been over 20 years since the first ranking of universities was published in the United States. Since then, numerous other efforts to rank or rate universities or university departments have appeared. University report cards have become both popular amongst the general public and increasingly important for academic institutions. A detailed examination of 10 report cards from around the world that evaluate and focus on either undergraduate choice or overall standing finds that the various schemes share broad principals and approaches but differ considerably in the details. Many of these variations can be attributed to differences in aims, higher education systems and cultures, and the availability and reliability of data.

Twenty Years of University Report Cards: Where are we now?*

The first ranking of universities was published in 1983 by *U.S. News & World Report*, an American periodical. Since then, efforts to rank and rate universities and university disciplines have flourished (see Jobbins (2002) for discussions of such work in the UK; Federkeil (2002) in Germany; Siwinski (2002) in Poland; Yonezawa, Nakatsui and Kobayashi (2002) in Japan; Filinov and Ruchkina (2002) in Russia; and Ledesma (2002) in Latin America). In 2002, it was estimated that over 20 such report cards existed (Kozmiński, Andrzej & Sadlak, 2002); by now that number is almost certainly much higher.ⁱ

Such undertakings have often received considerable criticism, particularly from academics, and particularly concerning methodological issues (Clarke, 2004; Gater, 2002; Crissey, 1997; Mallette, 1995; McGuire, 1995; Kersten, 2000; Eccles, 2002; Carrico, Hogan, and Athanassopoulos, 1997; Bowden, 2000; Goldstein and Spiegelhalter, 1996; Provan and Abercromby, 2000).ⁱⁱ Nevertheless, the huge popularity of these comparative evaluations indicates that they are clearly of great interest to consumers, and as a result, universities have been forced to pay attention to them.ⁱⁱⁱ Moreover, done carefully and rigorously, report cards provide important information to everyone interested and involved in higher education.

Much of the scholarly discussion on this topic has focused on the strengths and (more often) weaknesses of individual report cards (in particular, the venerable U.S. News

rankings) within single countries. Now that a critical mass of report cards exists and with increasing cross-national agreement as to what constitutes university excellence, the time seems right to conduct a survey of these report cards; to look for commonalities and differences and to see what we can learn about such exercises.

Report Cards Included^{iv}

As the number of report cards has multiplied, so too have the purposes of such undertakings. Originally focused on providing information to prospective undergraduates, report cards now evaluate graduate and professional schools (particularly MBA programs), provide measures of overall standing, and serve as government funding schemes.^v Moreover, whereas most of these report cards are national in scope, others are cross-national or even international.

In order to keep the exercise manageable and also to compare like with like, I will focus on report cards that meet the following two criteria: (1) they evaluate universities within either a single country or region; (2) they are reported in respected publications and include all or nearly all of the universities in that country or region. An additional advantage to these criteria is that they screen for the types of evaluations that have the broadest public appeal and have been most discussed among both the popular and academic press. The result is 10 report cards that focus on either undergraduate choice or overall standing, and which evaluate universities in five different individual countries and two geographical regions.^{vi} I will now compare these 10 schemes in some detail.

Table 1: Report Cards Included in Survey

Producer	Name of Report Card	County/Region	Stated Purpose
AsiaWeek*	Asia's Best Universities	Asia	Honouring excellence
The Center	The Top American Research Universities	United States	Identify top research universities
CHE/Stern	CHE and Stern University Rankings	Germany	Help school leavers decide where to study**
Good Guides	The Good Universities Guide	Australia	Undergraduate choice
The Guardian	University Guide	UK	Undergraduate choice
Macleans	University Rankings	Canada	Undergraduate choice
Melbourne Institute ^{vii}	Melbourne Institute Index of the International Standing of Australian Universities	Australia	International standing of Australian universities
Perspektywy	Table of Universities	Poland	Undergraduate choice
The Times	Good University Guide***	UK	Undergraduate choice – focus on teaching rather than research
U.S. News	America's Best Colleges	United States	Undergraduate choice

* AsiaWeek ceased publication as of December, 2001. Its online archives, including its university rankings, are still available.

** School leavers is a sub-category of prospective undergraduates; it includes only those students going directly from high school to university.

*** Note the similarity in names between report cards by Good Guides and The Times

Survey of Report Cards

Producers

Most of the report cards included in this survey are produced by magazines, newspapers, or book publishers.^{viii} While such origins presumably increase visibility, they may have also aroused suspicions among academics and others as to their theoretical and

methodological soundness. In response, a number of these publications have gone to great efforts to explain their methodology and to create advisory boards with whom they consult. Perhaps the best approach is that of CHE/Stern which combines the best of both worlds: the Center for Higher Educational Development, a non-profit consulting group, is responsible for its report card's concept and data, while the weekly news magazine, *Der Stern*, is responsible for marketing and distribution.

Table 2: Producers of Report Cards

Media Entity	Non-profit Research Institute	Combination
AsiaWeek	The Center	CHE/Stern
Good Guides	Melbourne Institute	
The Guardian		
Macleans		
Perspektywy		
The Times		
US News		

University-based versus Discipline-based

Considerable debate exists as to the appropriateness of evaluating universities as a whole versus by discipline. On the one hand, students may have to apply to a discipline rather than to a university, and quality within a university may vary by discipline so averaging across departments can produce a distorted view of the university's quality. On the other hand, variability across disciplines is diminishing, particularly among the best universities; students are not always certain which discipline they are most interested in when they investigate universities; and there is arguably an important “brand effect” for a university as a whole (see Toutkoushian, Dundar, & Becker (1998)). Certainly much

more work is involved to gather and analyse data by discipline, and issues of data validity and reliability are always a concern.

Most of the report cards included in this discussion evaluate universities as a whole. Two report on both universities and disciplines, and one (CHE/Stern) rates by discipline only.

Table 3: Whether Report Cards Rate Universities or Disciplines

	University as a whole	By discipline	Both
AsiaWeek	•		
The Center	•		
CHE/Stern		•	
Good Guides			•
The Guardian			•*
Macleans	•		
Melbourne Institute	•		
Perspektywy	•**		
The Times	•		
US News	•		

*The Guardian's ranking by university includes only a single indicator: the teaching assessment score.

** *Perspektywy* also publishes a *Prestige Ranking of Specialisation*, but it is based only on opinions of large employers and young professors.

Rankings versus Clusters

Some criticism has focused on the use of rankings to report results of evaluations, because differences between closely ranked institutions or disciplines are often due to statistical noise rather than true differences (see, for example, Clarke, M.(2002); Merisotis (2002)). As a result, some report cards have used clusters (i.e., high, medium, low) instead, listing entities within clusters alphabetically. This solution is not without its own methodological difficulties, however, as choice of cut-off points between clusters

may be rather arbitrary, and the same issue of true differences remains: the apparent distinction between the lowest-rated university and discipline in the “top” group, for example, and that of the highest-rated one in the next group may not be real. Moreover, clusters are simply not as satisfying as rankings. A possible solution is to rank but provide the consumer with easily understood information about the extent to which apparent differences in rankings reflect true statistical differences.

All but two of the report cards in this survey rank universities or disciplines rather than cluster them into groups.

Table 4: Whether Report Cards Report Rankings or Clusters

	Rankings	Clusters
AsiaWeek	•	
The Center	•	
CHE/Stern		•
Good Guides		•
The Guardian	•	
Macleans	•	
Melbourne Institute	•	
Perspektywy	•	
The Times	•	
US News	•	

Assignment of Weights

Another popular criticism of report cards, in addition to that regarding ranking, is the assignment by the producers of the report card of weights to the various measures included in the evaluation (NORC, 1997; Eccles, 2002; Carrico, et al., 1997; Clarke, 2002). The general argument is that choice of weights is subjective and arbitrary with little or no theoretical or empirical basis.

The difficulty, of course, is how to report results without assigning weights, since the various measures cannot then be combined into any overall ranking or clustering. One solution is to evaluate universities or disciplines separately on each measure, as is done by Good Guides.^{ix} Another is to allow the user to apply weights, as is done by CHE/Stern.^x An alternative response to some of the criticisms of applying weights is to survey others regarding what weights to apply, as is done by Melbourne Institute.^{xi}

Table 5: Whether Report Cards Assign Weights to Measures

	Weights Assigned	No Weights Assigned
AsiaWeek	●	
The Center	●	
CHE/Stern		●*
Good Guides		●**
The Guardian	●	
Macleans	●	
Melbourne Institute	●	
Perspektywy	●	
The Times	●	
US News	●	

* Examples are provided using five “key indicators” (presumably with equal weights applied).

** No weights are assigned to the 17 categories by which universities are evaluated; for categories that include more than one indicator, no explanation of weights is provided.

Indicators

The 10 report cards include a total of 72 different indicators spread across seven broad categories: Quality of Academic Staff/Faculty, Quality of Incoming Undergraduate Students, Quality of Undergraduate Program, Quality of Graduate Programs, Resources, Stakeholder Opinions, and Other (please refer to Appendix A for a complete table with detailed explanations of each indicator). Some of these indicators are included in more than one report card, but a considerable number are unique. No indicator is used by all

10; the closest is Score on national entrance exam/ high school performance (within the category, Quality of Incoming Undergraduate Students), which is used by eight of the 10.

Table 6: Broad Categories of Measures

	Quality of Academic Staff	Quality of Incoming Undergraduates	Quality of Undergraduate Program	Quality of Graduate Program	Resources	Stakeholder Opinion	Other
AsiaWeek	•	•			•	•	•
The Center	•	•		•	•		
CHE/Stern	•		•	•	•	•	•
Good Guides	•	•	•		•	•	•
The Guardian	•	•	•		•		
Macleans	•	•	•	•	•	•	•
Melbourne Institute	•	•	•	•	•	•	
Perspektywy	•	•		•	•	•	•
The Times	•	•	•		•		
US News	•	•	•		•	•	•

Quality of Academic Staff

All 10 evaluations include at least one indicator within the broad category, Quality of Academic Staff. Three of the 10, however include no measure of research output (The Guardian, Perspektywy, and US News); three use no measure of faculty prestige (CHE/Stern, The Guardian, and The Times); and seven of the ten do not include an indicator for teaching quality; the three that do are The Good Universities Guide, The

Guardian, and The Times (although CHE/Stern asks students their opinions about quality of teaching).^{xii} Only one scheme – The Good Universities Guide, includes at least one measure from each of the three sub-categories of Quality of Academic Staff: research, prestige, and teaching quality.

Table 7: Quality of Staff Measures Included

	<i>Research</i>	<i>Prestige</i>	<i>Teaching</i>
AsiaWeek	•	•	
The Center	•	•	
CHE/Stern	•		
Good Guides	•	•	•
The Guardian			•
Macleans	•	•	
Melbourne Institute	•	•	
Perspektywy		•	
The Times	•		•
US News		•	

Part of the explanation for this variation may be differences in foci. The Center, for example, with its emphasis on research, presumably sees no need to include a measure of teaching quality. The Guardian, in contrast, explicitly states that its focus is on teaching, not research; it is not surprising, therefore, that it does not include any measures of research output. CHE/Stern, however, states that good research is clearly related to good

teaching (Federkeil, 2002) and therefore includes a research income measure as well as opinions from students regarding teaching quality.

It should also be noted that obtaining independent, objective measures of teaching quality is difficult, expensive, and time-consuming. The UK, for example, spends a considerable portion of its higher education budget to conduct its Research Assessment Exercise which assesses the teaching quality of each discipline across universities. A less arduous, though arguably also less rigorous and reliable, measure of teaching quality is opinions of graduates, as is used by Good Guides and CHE/Stern.

In some respects, given the ease via the Internet with which one can gather such measures as publication counts, citations, and membership in academies, it is somewhat surprising that these indicators are not more popular across the report cards.^{xiii} Research grants and income, similarly, should be fairly easy to come by – universities must surely keep track of these numbers, yet half of the schemes do not include either of these indicators.

Quality of Incoming Undergraduate Students

Indicators used by the various report cards that fall into this category are:

- Score on national entrance exam/ high school performance
- Minimum cut-off scores required for entry
- University acceptance rate/ Demand for places
- Geographic diversity.

Much criticism exists regarding the inclusion of student intake measures (Clarke, 2002; Eccles, 2002), particularly intake scores. As an input variable (as opposed to a process or outcome variable), it is unaffected by quality of university. Macleans argues that smarter students results in a better learning environment, although others argue that there is no research to support such a claim. Moreover, they state that the inclusion of this measure is elitist – the students who do best on these tests tend to be the more advantaged. Finally, high schools vary widely in quality; therefore, the best high school grades do not necessarily equal the brightest students.

Nevertheless, such measures are popular across the board, with only CHE/Stern not including any indicator within this category, and this omission is only because German universities are not free to choose which students they admit (Federkeil, 2002:394). Indeed, as mentioned earlier, Student intake scores is the closest to a unanimous indicator as exists among these 10 report cards.

Quality of Undergraduate Program

Indicators include:

- Degree classification (honours, etc.)
- National academic awards won by students
- Graduates who go on to get Ph.D.s/ enrol in further study
- Graduation/ retention attrition rate
- Unemployment rate of graduates
- Average starting salary

- Student-faculty contact
- Contact between students
- Value-added score

Somewhat surprisingly, three of the report cards do not include any measures within this category (AsiaWeek, The Center, and Perspektywy). One could argue that quality of teaching (a measure used by The Guardian) really measures the quality of undergraduate program. Moreover, two report cards include measures of student opinion regarding their undergraduate education (CHE/Stern and Melbourne Institute), which are more subjective measures of undergraduate quality. AsiaWeek includes a measure of student acceptance rate (sometimes referred to as a university's "yield rate") which might be considered an indicator of undergraduate quality. Of the two remaining report cards – The Center and Perspektywy, neither is focused on undergraduate choice. Nevertheless, one might argue that even research universities must pay attention to the quality of their undergraduate program.

Of the indicators in this category, Graduation/ Retention/ Attrition rates is the most popular, with four of the score cards including at least one of these measures.

Quality of Graduate Programs

Indicators included across the report cards are:

- MAs/PhDs awarded
- International students

- Pass/fail rate on professional exams
- Preparation for post-degree career

Because many of the report cards are focused on prospective undergraduates rather than on post-graduates, it is not surprising that few of the evaluations include any measures of quality of graduate programs. One might argue that high quality graduate programs attract high quality faculty, but then one need only include a measure of quality of faculty. One might also argue that high quality graduate programs raise the overall academic level of the university, thus providing a more stimulating environment for undergraduates, although there appears to be little research to support this connection.

Of the five report cards that include measures of quality of graduate programs, two (The Center and Melbourne Institute) do not have as their primary focus undergraduate choice.

Resources: Student and Faculty Support

This broad category of measures contains the largest number and types of indicators across the 10 report cards. Evaluations vary tremendously in the number of such indicators they include, from a low of two by The Guardian and Melbourne Institute, to a high of 20 by CHE/Stern. No single measure is used by all evaluations; Student: Faculty ratio is the most popular, included in seven, while Total spending per student is used by five.

Indeed, the level of agreement across report cards as to which Resource measures should be included is generally low. For example, some schemes, such as Macleans, use a number of library measures – size, spending, and currency of collection, whereas The Center, Good Guides, The Guardian, Melbourne Institute, and US News ignore libraries altogether. AsiaWeek and Perspektywy include two Computers/IT measures, whereas only three other schemes include even one such indicator. Even at the level of sub-category, the various evaluations do not agree on which types of measures to include; no pair of report cards uses the same set of measures. All except The Center, however, do include at least one measure of Student support: academic.

Table 8: Resources Measures

	<i>Library</i>	<i>Computers /IT</i>	<i>Facilities/ Infrastructure</i>	<i>Student Support: Academic</i>	<i>Student Support: Non-Academic</i>	<i>Total Spending per Student</i>	<i>Staff Support</i>	<i>Revenue/ Assets</i>
AsiaWeek	•	•		•		•	•	
The Center							•	•
CHE/Stern	•	•	•	•	•		•	
Good Guides		•		•	•			•
The Guardian				•		•		
Macleans	•			•	•	•		•
Melbourne Institute				•		•		
Perspektywy	•	•	•	•	•			
The Times	•	•	•	•				
US News				•		•	•	•

These differences among the report cards regarding number and types of Resource measures included probably reflect, at least to some extent, differences in higher education systems and cultures and availability and reliability of data. Perceptions of excellence in Asian universities, for example, may have little to do with non-academic student resources, and thus AsiaWeek does not include any indicators in this category. A Seats: students ratio, used by Perspektywy, simply has no connection to university quality in most of the other systems. Variability amongst universities in levels of resources such as facilities/infrastructure and student academic support is likely greater in Germany than it is in, say, Canada or Australia.

Availability and reliability of data on university resources seems to be particularly problematic (Eccles, 2002). Individual universities often use different definitions of identical terms to compile their data, which can be very difficult to sort out. AsiaWeek has the added burden of working with data from different countries. Moreover, desired data is sometimes simply not available.

Stakeholder Opinion

This category is obviously not mutually exclusive of the others – stakeholders might be asked their opinions about quality of teaching, prestige of faculty, quality of undergraduate or graduate education, etc. Of the 10 report cards, seven include opinions of stakeholders. The group most commonly surveyed is academics – typically heads of universities or departmental Deans; five of the evaluations include such opinions. Three

ask employers for opinions, and three ask students. None gathers opinions from all three groups of interested parties.

The types of opinions gathered range from responses to a single question asking academics to rate each university on a particular scale, to responses to multiple questions asking graduates their opinions on a range of issues regarding their educational experiences. CHE/Stern includes a particularly creative question that asks professors where they would send their own children.

A major concern with including opinions of academics or employers asked to rate a long list of universities or departments is that they do not have enough specific knowledge about each and opinions are reflective of factors other than the criteria they are asked to use (Jacobs, 1999; Toutkoushian, et al., 1998).

Other Category

Indicators that do not fit into the other six categories include:

- Proportion of full-time faculty
- Undergraduate classes taught by tenured or tenure-track professors
- Student acceptance rate
- Number/ proportion of graduate students
- Diversity of student body (other than geographic)
- Number of required classes
- Class size

- Entry flexibility
- Comprehensiveness of life support services
- Time to degree
- Cost of living/ general information on city/town

In most cases, the number of indicators used by a given report card that does not fit into one of the six broad categories is relatively small. For Good Guides, however, the number is fairly substantial (seven of 26). The explanation appears to be the particular focus of this evaluation, which aims to broaden its appeal beyond the traditional, high-achieving student. It therefore includes a number of indicators of student diversity, as well as measures of entry flexibility, comprehensiveness of life support services, and cost of living/ general information about university city/town.

Some of the other indicators that fall into this category seem to be included by a report card because of particular features of the higher education system in that country. Time to degree, for example, is a particularly important factor in Germany, where the length of time it takes to attain a particular degree can differ by as much as two years depending on where one studies (Federkeil, 2002).

The only indicators in this category included in more than one report card are Class size, used by Macleans and US News, Number/ Proportion of graduate students, included by AsiaWeek and Perspektywy, and Cost of living/ general information about city/town, used by CHE/Stern and Good Guides.

Number of indicators

The number of indicators used by each report card, and their distribution across the seven broad categories varies enormously. Total numbers range from a low of six by The Guardian to a high of 33 by CHE/Stern.^{xiv} All of the evaluations include a relatively large number of Resource measures – this broad category comes in either first or second for each report card. Quality of Staff measures are also relatively numerous in most cases. Beyond these basic trends, however, there are few commonalities across report cards regarding number of indicators included.

Table 9: Number of Indicators within each Broad Category

	Quality of Academic Staff	Quality of Incoming Under-graduates	Quality of Under-graduate Program	Quality of Graduate Program	Resources	Stakeholder Opinion*	Other	TOTAL
AsiaWeek	6	2	0	0	8	2	2	20
The Center	4	1	0	1	3	0	0	9
CHE/Stern	1	0	2	5	20	2	3	33
Good Guides	5	3	3	0	7	1	7	26
The Guardian	1	1	2	0	2	0	0	6
Macleans	4	4	3	1	7	1	3	23
Melbourne Index	4	1	2	1	2	2	0	12
Perspektywy	2	0.5	0	0.5	10	2	2	17
The Times	2	1	3	0	3	0	0	9
US News	1	3	3	0	4	1	3	15
TOTAL	30	16.5	18	8.5	66	11	20	170

*I count type of respondent (academics, employers, or students) rather than number of questions asked. Therefore, the largest possible number of indicators in this category for each report card is three.

Size of Weights

Results regarding size of weights applied to the seven broad categories of measures (for the eight report cards that weight measures) are similar to those for numbers of indicators. In general, Quality of Academic Staff and Resources receive the highest weights, except the former nudges out the latter as the most important category overall. The report cards that do not follow this general pattern of ascribing their two highest weights to these two categories (regardless of rank order) are Perspektywy and US News, both of whom apply their highest weight to Stakeholder Opinion, and The Melbourne Institute, which places its second-highest weight on Quality of Graduate Programs.

Table 10: Weights (%) Applied to Broad Categories

	Quality of Academic Staff	Quality of Incoming Under-graduates	Quality of Under-graduate Program	Quality of Graduate Program	Resources	Stakeholder Opinion*	Other
AsiaWeek	23	14	0	0	32	20	10
The Center	33	11	11	11	33	0	0
CHE/Stern	n/a	n/a	n/a	n/a	n/a	n/a	n/a
Good Guides	n/a	n/a	n/a	n/a	n/a	n/a	n/a
The Guardian	40	10	15	0	20	0	15
Macleans	17	15	10	1	27	16	15
Melbourne Institute	40	11	14	16	11	8	0
Perspektywy	23	2.5	0	0	20	50	5
The Times	37	9	27	0	27	0	0
US News	3	16	23	0	23	25	11

When Quality of Staff is broken down into Research output/ Prestige and Teaching, however, once sees large differences among the report cards, with The Guardian placing all its weight within this category on teaching, The Times applying weights to both Research/ Prestige and Teaching, and the rest placing all their weights onto research/prestige.

Table 11: Weights (%) Applied to Quality of Academic Staff, Broken Down into Sub-categories

	Quality of Academic Staff: Research/Prestige	Quality of Academic Staff: Teaching
AsiaWeek	23	0
The Center	33	0
CHE/Stern	n/a	n/a
Good Guides	n/a	n/a
The Guardian	0	40
Macleans	17	0
Melbourne Institute	40	0
Perspektywy	23	0
The Times	14	23
US News	3	0

Disagreement also appears regarding what weight to place on stakeholder opinions – whether those of academics, employers who hire graduates, or the students themselves. Overall, stakeholder opinion is weighted from a low of 0 per cent by The Center and The Times, which do not include such measures, to a high of 50 per cent by Perspektywy. To some extent this difference mirrors a distinction between reputation and current performance. If one is concerned only with current performance, there may be no need to ask for opinions, which, especially when gathered from academics and employers, will result largely in a measure of reputation. On the other hand, one can also argue that reputation is a legitimate factor to include in assessing the quality of a university or discipline.

Discussion

Some 20 years of university report cards has resulted in schemes that share some broad principals and approaches but differ considerably in the details. Most are produced by media outlets and include rankings of universities using weights established by its authors. All approaches recognise the contribution of high quality of academic staff, good students, and substantial resources to university excellence, although decisions regarding whether the emphasis for quality of staff is on research versus teaching, and which resource indicators are included, differ considerably across schemes. Agreement also exists regarding the importance of including measures of quality of undergraduate program, although some of these measures, for example teaching quality, student-staff ratios, and opinions of stakeholders, also cross over into other categories.

Less agreement exists regarding the importance of quality of graduate programs, with about half of the report cards containing such measures. When the four schemes whose focus is broader than undergraduate choice are omitted from the analysis (AsiaWeek, The Center, Melbourne Institute, Perspektywy), greater consensus emerges to ignore graduate programs.^{xv}

Other variations among the report cards also seem to be explained by the somewhat different foci of the various evaluations, as well as by differences among higher education systems and cultures, and the availability and reliability of data. In some cases, however, it appears that different authors simply have different ideas about how to best measure university quality.

Below is an example of a league table constructed based on the averages of the 10 report cards included in this survey.

Table 12: Composite Report Card

Category	Number of Measures	Weight	Indicator	Number of Report Cards (out of 8) that Include this Indicator
Quality of Academic Staff	3	27%	Academic degrees held	5
			Research Income	5
			Teaching quality	3
Quality of Incoming Undergraduate Students	2	11%	Score on national exam/ High School performance	8
			University acceptance rate/ demand for places	3
Quality of Undergraduate Program	2	14%	Graduation/ retention/ attrition rate	4
			Unemployment rate of graduates	3
Quality of Graduate Programs	1	4%	MAs/PhDs awarded	3
Resources	5*	24%	Student:Faculty ratio	7
			Total spending per student	5
			Annual giving	4
			Library: size	3
			Library: spending	3
Stakeholder Opinion	1 (academics)	15%	Academics (one question asking them to rate universities on a given scale)	6***
Other	2	7%	Class size	2
			Cost of living/ general information on city/town	2
TOTAL	16**	100%****		

*Average without CHE/Stern, which includes 20 indicators in this category, a clear outlier

**Actual total is 17 with CHE/Stern Resources count included

***Four of the five ask this kind of question.

**** Total percents actually add up to 102 because of rounding.

Conclusion

University report cards are not only tremendously popular among students and parents, but also are increasingly accepted (if grudgingly in some cases) by academia as useful tools for gauging and tracking academic performance and reputation. Criticisms have forced the creators of these tables to more exactly specify their aims, improve their methodologies, consult with Advisory Boards, and increase the transparency of their undertakings. In addition, universities and governments have been nudged into improving the quality and reliability of the higher education data they collect.

First published in North America, such rankings have spread to Europe and the Asia-Pacific region, and are under discussion or in the early stages in such countries and geographic regions as Russia, Latin America, and Nigeria, among others.

Increasingly, users of report cards can, via the Internet, choose the indicators of greatest interest to them and ascribe their own weights or priorities. Such an approach provides a solution to two of the most persistent criticisms of report cards – that weights are subjective and arbitrary, and that the evaluations are largely geared towards a narrow group of users.

Another clear trend is towards global rankings of universities, which has already been attempted by SJTU and The Times HES. As universities increasingly compete in a global environment, students and faculty continue to look beyond their own country's

borders, and cross-national agreement increases as to what constitutes high quality universities, such an exercise becomes not only feasible but also important and necessary.

End Notes

-
- * I would like to thank Ross Williams for his invaluable thoughts on the initial conceptualisation of this article and helpful comments on earlier drafts; I would also like to thank Dean Goodman for several good suggestions and also for his help and understanding in providing me the time to complete this article.
- i Gormley and Weimar (1999) and refer collectively to published comparative evaluations of universities as report cards.
- ii Some Deans upset with the law school ratings by U.S. News & World Report have launched an anti-ranking campaign (Carter, T. (1998).
- iii See Cohen, David (1999) for a discussion of reactions to university rankings in Asia. One reason for this criticism, in addition to methodological problems and general defensiveness of academia, was the use of the term, “best,” in many of the ranking schemes. One can argue that “best” is in the eye of the beholder, or, in this case, the student, and that to state that a certain number of indicators added together equaled “best” was simply indefensible.
- iv Research for this article was begun in mid-2004; therefore, the rankings I discuss are mostly those from 2003. All information regarding the details of these report cards is from the following sources, unless otherwise noted: AsiaWeek: <http://www.asiaweek.com/asiaweek/features/universities2000/>; TheCenter: <http://thecenter.ufl.edu/>; CHE and Stern University Ranking: <http://www.daad.de/deutschland/en/2.2.9.html>; Federkeil, G. (2002); The Good Universities Guide: <http://www.thegoodguides.com.au/>; The Guardian: <http://education.guardian.co.uk/universityguide2003>; Macleans: <http://www.macleans.ca/universities/TheMelbourneInstitute>: <http://www.melbourneinstitute.com/austuniv/austuniv.html>; Perspektywy: Siwinski, W. (2002); The Times: http://www.timesonline.co.uk/section/0_716.00.html (access by subscription only); U.S. News & World Report: http://www.usnews.com/usnews/edu/college/rankings/tankindex_brief.php.
- v A few examples are Gourman Report: Graduate Programs (graduate and professional programs in the U.S.), National Research Council (U.S.) Committee for the Study of Research-Doctorate Programs in the United States, US News & World Report. America’s Best Graduate Schools, Australian Financial Review Boss (MBA programs in Australia), B-School Net (a site ranking business schools and MBA programs in Germany), Canadian Business (MBA programs in Canada), Financial Times (MBA programs throughout the world), Marr/Kirkwood Side by Side Comparison of International Business School Rankings, MBA Rankings – Pforzheim (world’s top 50 business schools), New Zealand Performance-Based Research Fund, Higher Education and Research Opportunities in the UK (Research Assessment Exercise conducted by the four higher education funding bodies in the UK).
- vi Australia, the United States, and the UK are each represented by two different report cards.
- vii The author is also one of the producers of the Melbourne Index.
- viii For purposes of clarity and brevity, I will refer to each of the report cards by the name of its producer, or in some cases a truncation of this name.
- ix Several of the measures include more than one indicator, and for these there is no explanation of weights applied to these indicators.
- x CHE/Stern also publishes “selected results” in which it chooses five indicators (which are presumably given equal weights) and presents results.
- xi We surveyed non-Australian heads of universities and Deans of Australian universities
- xii I have included opinion surveys in a separate broad category: Stakeholder Opinions.
- xiii Caution, however, must be exercised in using ISI data and citation counts. van Raan (2004) discusses various problems with the use of this data. Van Leeuwen, et al. (2001) show that English-language articles are cited at a higher rate than are non-English language articles. As a result, if a university publishes a lot in a non-English language, it is doubly disadvantaged – first, because fewer non-English journals are included, and second, because even for those included, their articles are less frequently cited.
- xiv But not all at once; each discipline includes its own set of indicators; my own playing around resulted in

about 25 indicators per discipline. Nevertheless, such a number still represents the high end of the scale.

xv AsiaWeek does not include any measures of quality of graduate program.