University Rankings 2.0

New frontiers in institutional comparisons

Alex Usher

Educational Policy Institute, Toronto, Canada

The number of university rankings systems in use around the world has increased dramatically over the last decade. As they have spread, they have mutated; no longer are ranking systems simply clones of the original ranking systems such as *US News* and *World Report.* A number of different types of 'mutation' have occurred, so that there are now varieties of rankings around the world. The purpose of this short paper is to describe these mutations and examine likely future developments in rankings as they continue to spread across the globe.

University rankings are not a new phenomenon. In fact, they date back to the beginning of the 20th century, when some US states began publishing institutional pass rates on state licensing exams in things such as law and dentistry. Later, Henry Herbert Maclean worked on a series of so-called 'genius studies'. The first, entitled 'Where We Get Our Best Men', provided statistics on the nationality and educational background of the country's most prominent scientists and men of business – the high counts for places like Harvard and Yale were taken as proof that these institutions were the country's best. Other early attempts to classify and rank institutions involved interviews of institutional officials such as Presidents, Deans or Department Heads, either asking them what they thought of the quality of graduates of various institutions (in the case of Kendrick Babcock's work on behalf of the US Bureau of Education and later the Association of American Universities) or asking them who they thought the 'best men' in their respective disciplines were and then developing rankings based on the number of 'best men' who matriculated at various institutions (a similar logic is at work today in the Shanghai Jiao Tong rankings' use of alumni Nobel prizes and Field medals as an indicator).

Remarkably, the top ten institutions in these ranking from over a hundred years ago looks very similar to the top ten in current rankings such as *US News* and *World Report.*

In the 1960s, with the development of large scientific databases such as the Science Citation Index and the Social Science Citation Index, it became possible to provide some quantitative measurements of academic staff members' output, and various journal articles appeared comparing these. Indeed, these statistics also played some role in the 1982 Assessment of Research Doctorate Programs conducted by the US National Academy of Sciences.

These rankings, perhaps because of their scientific and quantitative nature and the fact that they only purported to rank graduate programmes, did not provoke much controversy. It was only in the mid-1980s, when US News and World Report began ranking entire universities and, more specifically, touted ranking as a tool to assist in the selection of undergraduate institutions, that real controversy was aroused and people began to view rankings as a dangerously reductionist way of evaluating education. Yet, its reductionist character was also part of what intrigued the public about rankings: they appeared to illuminate certain aspect of institutional quality which had previously appeared opaque. And, as tuition fees began to appear in new countries (such as in the UK and China in the late 1990s) or increase rapidly (such as Canada in the early

1990s), the need for consumer guides to evaluate educational investments grew, and rankings seemed to fit that bill rather nicely.

These original 'classic' rankings – such as US News and World Report and other closely modelled on it like Canada's Maclean's rankings or Poland's Rzeszpospolita – essentially shared seven key features. In North America, these seven features are often believed to be intrinsic to rankings even though (as we shall see) rankings that violate each of the seven attributes exist. These features of classic rankings are:

- They focus on the undergraduate experience and are intended as tools to help guide students and parents choose between institutions. Their choice of indicators is thus made with this end in mind.
- They are national in scope, dealing with a single domestic education market.
- They compared entire institutions. That is to say that the units being compared were entire institutions rather than smaller units such as faculties and departments.
- Rankings were done on an ordinal scale, arrived at using scored indicators which were aggregated and summed.
- Data and rankings were presented so as to present a single story; there could be only one 'winner':
- Data tended to come from either 'official' government sources or surveys of institutions themselves.
- The process of ranking was managed by commercial media outlets.

However, as rankings have spread around the world, a number of different rankings efforts have managed to violate every single one of these principles.

The first major area where these principles were breached was with respect to rankings being solely about undergraduate education. Among the most famous rankings on the world now are the Shanghai Jiao Tong's Academic Ranking of World Universities (ARWU), where the indicators are almost exclusively concerned with research. Indeed, a number of rankings, particularly in Asia, are now largely concerned with research performance and are not properly speaking dealing with issues of undergraduate quality.

Closely related to this was the issue of doing international rankings, which was first done by the magazine Asiaweek in the late 1990s, when it tried to rank universities across Asia. More recently, both the ARWU and the Quacquarelli Symonds (QS)-Times Higher Education Supplement (THES) rankings have also provided international comparisons as well. International rankings almost by definition are more likely than national rankings to rely on research metrics for indicators. This is because institutions in different countries collect data in very different ways; as a result, bibliometrics are in effect the only internationally comparable metric available.

Across most of Europe, rankings are now available which compare departments rather than whole institutions. The Netherlands' Keuzegids Hoger Onderwijs and Elsevier rankings, the UK's Guardian and Italy's La Repubblica rankings are all examples of this phenomenon. In effect, these rankings disaggregate institutions to their constituent parts (a process which many within the academy believe is a much more valid form of comparison). These same European rankings also do away with the process of weighting individual indicators; the results of each indicator are presented separately, though most continue to show the schools (or departments) with the best scores across all indicators at the top. In a couple of cases, however – most notably Germany's CHE rankings – the rankers go one step further and do away with the concept of even presenting 'top' institutions. Instead, by using the interactivity of the web and liberating themselves from the newspaper or magazine format's requirement to tell a single story, they allow users to rank institutions based on their own choice of indicators (these rankings are sometimes called 'personalised' rankings, or 'do-it-yourself' rankings).

Another recent innovation in rankings is the increased use of survey data. While surveys of educators and employers have long played a role in obtaining data for reputational rankings, only recently have surveys of students and their views of their schools and their educational experiences begun to play a role. Germany's CHE rankings and Canada's Globe and Mail rankings both have a number of indicators which are populated by student survey data, as do the Dutch and Italian rankings noted above. This appears to show some promise in developing those ranking systems that wish to provide information to students using rankings to choose between undergraduate institutions because of the way they can provide real information about what institutions are really like.

The final and perhaps most interesting recent innovation in rankings is their adoption as a policy instrument by governments or government agencies in many countries. In a number of countries – Taiwan, Nigeria, Kazakhstan and Pakistan to name but a few – rankings are now being published by governmental or para-

governmental agencies as a tool to encourage institutions to strive for excellence. This fundamental change in the nature of rankings is more art than science, and one that many rankers themselves find somewhat troubling, not least because they themselves understand the limitations of the data and the way that weighting and aggregating indicators.

Though rankings have angered many, they seem set to continue to spread around the globe because they represent a convenient heuristic device for making the massive complexities of the university enterprise understandable.   As time goes on, however, there is an increasing understanding that rankings – at least those of the sort where indicator scores are weighted and aggregated to produce a single overall score and hence a sort of 'league table' – are essentially limited in that the choice of indicators and weights imposes a single definition of institutional quality. Since educational quality is really in the eye of the beholder and there are many possible definitions of quality, any single set of rankings will inevitably do an injustice to other definitions of quality. This does not necessarily mean that rankings are invalid – rather it means that multiple sets of rankings are required to pick up multiple definitions of quality.

The problem at the moment is that where we do find multiple rankings, they tend to show similar results – at the top at least. With global rankings now on the scene, most countries now have at least three different observations on their country's institutional performance. At the very top, they all tend to show the same thing – Harvard, Stanford and Yale are invariably top in the United States, as are Oxford and Cambridge in the United Kingdom, Toronto and McGill in Canada, and Beijing and Tsinghua in China. Where they disagree is further down the table – it is rare, for instance, that there is unanimity about which is the fifth-best university in a country. What this suggests is that the various rankings out there now are probably not measuring what they think they are measuring. Regardless of what indicators they select, most seem to be indirectly measuring some combination of institutional age (it is rare that a country's oldest institutions are not among its highest ranked), institutional size and financial clout. In other words, inputs.

The challenge, then, is to find other sets of indicators that can measure throughputs and value added in a more systematic way. This brings us to the question of data quality and data gathering. One of the most important things to understand about rankings is that their authors are fundamentally constrained by data collection. In many places, data on what universities actually do simply isn't very good, or is not collected in a consistent way across institutions. As a result, they tend to gravitate to the pieces of information that are easiest to collect, namely: inputs (student marks, finances and academic staff), research outputs (bibliometrics) and reputational surveys. Of these three, only bibliometrics really works on an international basis. Inputs are almost impossible to collect on a transnational basis and reputational surveys are bedevilled by problems of survey response rates (though this hasn't stopped Quacquarelli Symonds gamely trying these on). What they tend to ignore are serious aspects of the student experience such as teaching and institutional service missions.

It is for this reason that the emerging practice of using student surveys in ranking seems likely to catch on. By asking questions about student satisfaction, student experiences and student engagement, one can get reasonably comparable data about the general learning environment at different institutions. This applies to both national and international comparisons: the growth of Germany's CHE approach (it now runs similar rankings in both Switzerland and the Netherlands) seems to point to the possibility that international rankings might in time be able to transcend mere bibliometrics and provide a degree of multi-dimensionality which has hitherto been lacking in many rankings. No doubt, as time goes on, the limits and drawbacks of this approach will become more apparent. It is not clear, for instance, that all students enter with similar expectations about the quality of university services and this may systematically distort any rankings based on satisfaction. It is also not clear that the results of surveys on teacher satisfaction in North America and Europe are likely to be comparable to those in Asia, where teachers are generally accorded much greater respect. Nevertheless, as international rankings proliferate this seems certain to be a trend to watch, and the recent decision of the European Union to proceed with a pan-European ranking based largely on the CHE model makes it even more likely that this approach will spread.

The other possible significant development in rankings in the near-term is the emergence of some international standards in the reporting of institutional data. Though QS only reports on six indicators in its rankings, it has quietly been collecting data on a number of other indicators in its annual institutional survey to

see if it is in fact possible to harmonise certain data definitions (for instance, on volumes held in libraries). Should QS succeed in developing some kind of acceptable standard for reporting this kind of data, one would expect institutions around the world to adopt it fairly quickly as not only would it creep into rankings, but it would also provide institutions with some benchmarks that they currently lack. In the developing world at least, this would almost certainly be met with eagerness, at least by those institutions that have pretensions of joining the global elite.

A final point is that rankings are almost certain to continue spreading at a very rapid pace in the developing world. In developing countries, rankings are seen as beneficial for two main reasons:

1. They can encourage institutional transparency and create a culture of quality measurement in education. Higher education the world over has transparency issues, but this effect is multiplied in developing countries where nothing like a system of institutional research yet exists. But with no transparency, how can institutions be expected to improve? Rankings are not the only possible way to improve this situation, but they can play a role in changing institutional culture around self-assessment and data collection.

2. They can act as a spur to improved institutional performance. In more market-driven systems, rankings are often accused of being a leading force in the 'marketisation' of higher education. In countries like Vietnam or Kazakhstan, where market forces in higher education are weak, this is precisely why governments like the idea of rankings. In the absence of market forces, only techniques such as rankings – which has a kind of 'name and shame' aspect as far as poor performers are concerned – can get institutions to pay serious attention to remedying perceived lags in performance.

It has by now perhaps become trite to observe that 'university rankings are here to stay'. But what is clear from this short survey is that not only are they certain to stay - they are also going to evolve. Already, we have seen tremendous mutations in terms of rankings' purposes, methods of data collection, methods of data display, and choices of indicator. There is no reason to think that the innovation has yet stopped; indeed it is perhaps just beginning.

Alex Usher is the Vice-President (Research) and Director (Canada) of the Educational Policy Institute (EPI), Toronto, Canada