University World News

Can we measure education quality in global rankings?

Philip G Altbach and Ellen Hazelkorn14 August 2018 Issue No:517

The most influential global academic rankings – the highly influential Shanghai Academic Ranking of World Universities (ARWU), *Times Higher Education* (*THE*) World University Rankings and QS World University Rankings – have been in existence for more than a decade and are now a major force in shaping higher education worldwide.

One of their key purposes is to demonstrate the world's best universities, based on their own criteria. However, they consider fewer than 5% of the more than 25,000 academic institutions worldwide. The rankings are influential – based on rankings students make decisions on where to study; some governments allocate funds; and universities struggle to improve their position in them.

From the beginning, these rankings have focused primarily on research productivity. Reputational measures are also included in the QS and *THE* rankings, but they remain controversial due to low response rates that accentuate biases and limited perspective.

Each survey indicator is considered independently, whereas multi-collinearity is more persuasive – in other words, doctoral students, citations, research income, internationalisation, etc, are highly interdependent.

Allowing for some overlap, research-related indicators constitute approximately 70% of the total score for QS while reputation constitutes 50%. Both ARWU and *THE* are 100% based on research or research-related indicators.

Teaching-learning enters the rankings equation

Without question, teaching is the fundamental mission of most higher education institutions; with few exceptions, undergraduates comprise the majority of students enrolled in higher education worldwide. However, the 'world-class' concept is derived from those universities that score highest in global rankings. This is relatively easy to explain.

Research-intensive universities tend to be the best known internationally and hence, the most recognisable in reputational surveys. Bibliometric data is easily captured, albeit that practice continues to undervalue arts, humanities and social sciences research as well as research with a regional or national orientation – especially research published in languages other than English.

Global rankings have been quick to capitalise on finding a solution to this issue by including more indicators about the quality of education and teaching. Richard

Holmes pointed out that this remains "unmapped territory". However, the problem is more fundamental than the choice of indicators.

One reason teaching and learning has not been included in global rankings is the difficulty of measuring and comparing results across diverse countries, institutions and students.

In addition, there is the necessity to take account of how and what students learn and how they change as a result of their academic experience without simply reflecting the student's prior experience – their social capital. The focus is the quality of the learning environment and learning gain rather than the status or reputation of the institution.

Thus, many individual colleges and universities seek to assess teaching quality using a variety of measures, including teaching portfolios and peer-assessment, for the purposes of recruitment and promotion of faculty members. In many countries, faculty must acquire a credential in teaching and learning practice prior to, or upon, appointment.

More importantly, it is misplaced to think we can measure teaching, at scale, in a way that is distinct from the outcomes of learning. The concept of teaching quality as an institutional attribute is also problematic because research shows most differences occur within, rather than between, institutions.

Measuring education quality and student learning

The debate about educational quality takes different forms in each country, but increasing emphasis is being put on learning outcomes, graduate attributes, life-sustaining skills and, crucially, what higher education institutions are contributing – or not – to student learning.

In 2011, following the success of PISA (the Programme for International Student Assessment), the OECD piloted its Assessment of Higher Education Learning Outcomes (AHELO) project. By administering a common test to students in 17 countries, the aim was to identify and measure both good teaching and learning.

Developed to challenge the prominence of global rankings based primarily on research output, AHELO proved controversial and was suspended. PIAAC, the OECD Programme for the International Assessment of Adult Competencies, measures adults' proficiency in literacy, numeracy and problem-solving in technology-rich environments – and was first published in 2013.

Measures of teaching quality are being developed in several nations. In 2016, England pioneered the Teaching Excellence Framework (TEF). The initial government concept was controversial, not least because results were to be tied to funding. TEF was developed by a consortium of key stakeholders to assess undergraduate provision and will be extended to disciplinary (subject) level beginning in 2020.

National testing is another method: Brazil's Exame Nacional de Desempenho de Estudantes (ENADE) or National Examination on Student Performance assesses student competence in various professional areas. The exam is aimed at evaluating university

programmes and not student or academic knowledge. Likewise, Colombia has developed Saber Pro with similar objectives.

In the United States, the Collegiate Assessment of Academic Proficiency (CAAP), the Collegiate Learning Assessment (CLA) and the ETS Proficiency Profile seek to measure learning using national tests. There are also student self-reporting exercises, such as the National Survey of Student Engagement (NSSE) and, for the community college sector, the Community College Survey of Student Engagement (CCSSE).

NSSE assesses the amount of time and effort students put into their studies and other educationally relevant activities and how an institution deploys its resources and organises the curriculum. The NSSE programme has been duplicated in Australia, Canada, China, Ireland, New Zealand and South Africa with similar initiatives in Japan, South Korea and Mexico.

What global rankings are doing

All global rankings, including the European Union's U-Multirank, include indicators for educational quality – some more successfully than others:

- QS, *THE* and U-Multirank (the latter at discipline level) use faculty-student ratio. However, due to different methods by which faculty and students are classified between disciplines and within institutions and countries, this is considered a highly unreliable indicator of educational quality.

- QS, and *THE* include a peer survey of teaching, but it is unclear on what basis anyone can evaluate someone else's teaching without being in their classroom.

- ARWU uses Nobel Prizes or Field Medals awarded to alumni and faculty as a proxy for educational quality – which is clearly ridiculous.

THE has just launched its 'Europe Teaching Rankings', drawing on the experience of the *Wall Street Journal/Times Higher Education*College Rankings. Fifty per cent of that ranking is based on its own student survey and another 10% is drawn from its academic reputational survey. It also allocates 7.5% of the final score to the number of papers published and 7.5% to the faculty-student ratio.

The student surveys appear to draw from the American NSSE methodology, but there is considerable debate about the use of such surveys on an international comparative basis without ensuring a representative sample and accounting for differences among students and the shortcomings of self-reported data.

The *THE* surveys also use the proportion of female students (10%) as a measure of inclusivity, but this is questionable, given that female students accounted for 54.1% of all tertiary students in the EU-28 as of 2015.

Thus, it is worth noting how few underlying measures have anything to do with actual teaching – even if it is defined broadly.

Conclusion

Despite some scepticism about the methodological and practical aspects of a global methodology, the race is on to establish one. There are various actions by ranking organisations, governments and researchers to identify more appropriate ways, using more reliable data, to measure and compare education outcomes, graduate employability, university-society engagement, etc.

In a globalised world with mobile students, graduates and professionals, we need better information on how to evaluate an individual's capabilities and competencies.

But one of the lessons of rankings is that, without due care, indicators can lead to unintended consequences. We know that student outcomes will determine future opportunities. But conclusions based on simplistic methodologies could further disadvantage students who could and should benefit most if universities become more selective and focus on students most likely to succeed in order to improve their position in global rankings.

Thus, it is clear that creating reliable international comparisons of educational outcomes is extremely challenging.

Clearly, assessing teaching and learning is central to determining the quality of higher education, but using current methodologies to produce comparative data is foolhardy at best. Rather than fooling ourselves by believing that rankings provide a meaningful measure of education quality, we should acknowledge that they simply use inadequate indicators for commercial convenience.

Or, better yet, we could admit, for now at least, that it is impossible to adequately assess education quality for purposes of international comparisons.

Philip G Altbach is research professor and founding director of the Center for International Higher Education, Boston College, United States. Email: altbach@bc.edu. Ellen Hazelkorn is emerita professor and director, Higher Education Policy Research Unit, Dublin Institute of Technology, Ireland, and partner, BH Associates, education consultants. Email: ellen.hazelkorn@dit.ie.

This article was altered on 27 August to remove U-Multirank from the list of rankings that include peer survey of teaching, which is incorrect.