



STI 2018 Leiden

*23rd International Conference on Science and Technology Indicators
"Science, Technology and Innovation Indicators in Transition"*

STI 2018 Conference Proceedings

Proceedings of the 23rd International Conference on Science and Technology Indicators

All papers published in this conference proceedings have been peer reviewed through a peer review process administered by the proceedings Editors. Reviews were conducted by expert referees to the professional and scientific standards expected of a conference proceedings.

Chair of the Conference

Paul Wouters

Scientific Editors

Rodrigo Costas
Thomas Franssen
Alfredo Yegros-Yegros

Layout

Andrea Reyes Elizondo
Suze van der Luijt-Jansen

The articles of this collection can be accessed at <https://hdl.handle.net/1887/64521>

ISBN: 978-90-9031204-0

© of the text: the authors

© 2018 Centre for Science and Technology Studies (CWTS), Leiden University, The Netherlands



This ARTICLE is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

Why the public nature of university rankings torpedoes their value for benchmarking: an institutional researcher’s opinion

Cathelijnn J.F. Waaijer*

* c.j.f.waaijer@bb.leidenuniv.nl

Institutional Research Unit, Administration and Central Services, Leiden University, P.O. Box 9500, Leiden, 2300 RA (Netherlands)

Introduction

University rankings are here to stay. Since the launch of the Academic Ranking of World Universities (ARWU) ranking in 2003, a multitude of other international university rankings have been launched. Notable ones include the Quacquarelli Symonds (QS) World University Rankings, the Times Higher Education World University Rankings, the Leiden Ranking, and U-Multirank.

The surge of international university rankings over the past fifteen years has influenced the management of universities (e.g., Hazelkorn, 2011; Stolker, 2014). Several functions in universities deal with university rankings. These include university leaders, research policymakers, accreditation and quality assurance specialists, and marketing specialists. In addition, a specific function is often involved in supplying data to university rankers and analysing the outcomes of rankings: institutional researchers (Bryant, 2017).

Institutional researchers aid “the collection, analysis, interpretation, and communication of data, and the strategic use of information for effective decision making and planning” in universities (Association for Institutional Research, 2018). They do so in a multitude of areas, e.g., strategic planning, student retention, and alumni placement (Prewitt-Freilino & Rush, 2017). Benchmarking is an indispensable method to support effective decision making and planning, because it shows where an institution is doing well compared to others and where it could improve itself (Hazelkorn, 2011; Levy & Ronco, 2012).

The public nature of university rankings means these rankings have the potential to serve as a rich information source for benchmarking. However, in this paper, I will argue that it is the very fact that university rankings are public that makes it impossible for institutional researchers to use these rankings for their primary task: supporting university leaders in effective decision making and planning.

Fundamental issues with university rankings

The objectives of these university rankings vary, but all rankings have in common that they quantify universities’ performance by operationalizing this performance into indicators. For example, Times Higher Education describes its World University Rankings as “the only global university performance table to judge research-intensive universities across all of their core missions: teaching, research, knowledge transfer and international outlook” (Times Higher Education, 2018). The producers of the Leiden Ranking and U-Multirank describe their rankings as “offer[ing] key insights into the scientific performance of over 900 major

universities worldwide” (Leiden Ranking, 2017) and “comparing the performances of higher education institutions – in short: universities – in the five broad dimensions of university activity: (1) teaching and learning, (2) research, (3) knowledge transfer, (4) international orientation and (5) regional engagement” (U-Multirank, 2018), respectively. QS, finally, advertises its rankings by inviting its website visitors to “compare the world’s top universities with the latest edition of the QS World University Rankings®, and explore leading institutions by region and subject” (QS, 2018a).

How rankers determine universities’ performance varies to a great extent (Moed, 2017, see its appendix for a good overview of the used indicators and their weights). Several points of criticism are made on the suitability of rankings to determine universities’ performance:

1. The inability of a collection of indicators to together form a complete overview of universities’ performance; the complexity of reality cannot be reduced to a few indicators.
2. Related to the first point: several of the main university rankings (e.g., ARWU, QS, THE) apply weights to their indicators to get a composite score. As also indicated by other studies (e.g., Marginson & van der Wende, 2007), these weights are quite arbitrary. How valid are the weights that determine the composite measure and thereby the ranking of universities?
3. The validity of the indicators themselves to measure what they aim to measure. For example, does THE’s survey of teaching reputation really measure teaching reputation, or rather research reputation, to which it is highly correlated (Usher, 2013)?
4. The inability of numbers to reflect differences between institutions; what does a numerical difference mean? This question is related to the manner in which raw measurements are transformed to scores. The main rankings transform either by normalizing by the maximum, by computing percentile rank or by computing distance to the medium (Moed, 2017). What, then, does a difference of ‘1.0’ in a score mean?

Finally, several of the main rankings suffer from a lack of transparency, both on data and methodology (Hazelkorn, 2011; Marginson & van der Wende, 2007; Vernon, Balas & Momani, 2018). This is not a fundamental point of criticism as to whether rankings can determine universities’ performance, like the points above are. Still, it is an important point when it comes to benchmarking.

Benchmarking in higher education

For a university it is valuable to know how well it is doing in different aspects compared to other universities. What is the university’s quality of teaching, how well are the university’s programmes providing students with the knowledge and skills they need on the labour market, and how much income is the university generating? Traditionally, benchmarking data on education have been relatively plenty. In the United States, a multitude of data on postsecondary education is available through the Integrated Postsecondary Education Data System. Higher education institutions provide their own data to the National Center for Education Statistics (NCES), which then provides data from all institutions publically. In the Netherlands, data on education are compiled in the *1 cijfer HO* registry, which provides data on student enrolment, study performance, and study duration

Such data often get to be collected through a push from governments, more specifically through the demands of accreditation and quality assurance bodies (Haskell, 2017). Haskell notices an increase of the number, type and complexity of reports. At the same time, the

influence of accreditation and quality assurance bodies means a structured effort is made to collect and harmonize data, which helps to consistently monitor educational outcomes.

For research, such structured efforts exist to a much smaller degree. While quality assurance of research and research assessment are given much attention, this attention has not led to many structured efforts to collect and harmonize data on the research activities of universities. Of course, a large part of the reason for the lack of harmonization lies in the inappropriateness of metrics to fully grasp the complex reality of research quality (see Wilsdon et al., 2015, as well as the list of objections to university rankings in the section above). At the same time, the objections against using metrics in research assessment (such as indicators not measuring what they intend to) also apply to using metrics in teaching assessment (Shin & Toutkoushian, 2011). In that light, it is surprising that structured efforts for collection, harmonization and sharing of data on research trail so far behind those on teaching.

Here, some further reflection is warranted. Can benchmarks of research performance, and by extension research metrics, help university leaders to make better-informed decisions? Indeed, a rich set of literature exists that describes the negative effects of research metrics, whether intended or unintended (de Rijcke et al., 2016). At the same time, studying which factors shape an outcome helps to put the outcome in question into perspective. For example, in the realm of education there is a wealth of evidence showing which factors – academic, socio-economic and personal – influence student performance (e.g., McKenzie & Schweitzer, 2001). Through benchmarking, higher education institutions can not only compare themselves to similar institutions, but also see how dissimilar they are to other institutions with regards to the factors that shape performance. This also gives an estimation of the effects certain changes would have. Whether such changes would be desirable from the point of view of the institution's mission is of course a crucial question; an institution educating many students from poor backgrounds could have “better” student performance if it attracted more students from rich backgrounds, but such a change would not fit the institution's mission. The same goes for research performance: by benchmarking with a rich set of factors that help to explain research performance, understanding of the factors underlying research performance can be improved and institutions could make more conscious choices.

Benchmarking and university rankings

University rankings now are the main manifestations of universities' performance in research being compared. As indicated in the section “Objectives of university rankings”, many of these rankings are presented as lists that reflect university's performance across several dimensions. As such, rankings publish data on the performance of multiple universities. This very fact makes rankings interesting for university leaders, as the rankings indicate how well their own universities are performing compared to other universities – and can be used for benchmarking purposes. Compared to a world where no such information is available, this is an improvement. Of course, this is only the case when the published data are transparent, meaningful and fit the concept they aim to measure. A first problem that occurs here is the fact that the main outcome of many university rankings is a composite score. This score is often the result of non-transparent raw data (e.g., numbers of publications and citations, results from reputation surveys), non-transparent transformations to scores, and the application of weights. This makes it difficult to analyse which meaningful differences in performance exist between universities. Does a difference of ‘0.5’ in a score mean a university is doing much worse than another, or is it the result of, say, only a 0.1% difference in the raw measure?

Rankings that stand out positively with regard to non-composite measures are the Leiden Ranking and U-Multirank. These make it easier to use the rankings for benchmarking purposes than rankings that have composite scores. The Leiden Ranking aims to provide information on the scientific performance of universities. It is a ranking that is used often by institutional researchers, and even as a predictor of future scores on bibliometric indicators in other university rankings (personal communications EduData Summit, 12-14 June 2017). The Leiden Ranking also stands out in the fact it makes its raw data public (Vernon, Balas & Momani, 2018). Although the Leiden Ranking is valuable to institutional researchers and policymakers in higher education *exactly because* it aims to measure scientific performance (including collaboration statistics) using bibliometric indicators and limits itself to that aim, this is also a drawback for institutional researchers and university policymakers. For them, data on teaching, income, regional ties, and many more topics, are vital as well.

This is where U-Multirank could provide value. U-Multirank was first published in 2014, after an earlier initial feasibility study, with funding from the European Commission (U-Multirank, 2018). It provides data on a large variety of indicators, covering the topics education, research, knowledge transfer, international orientation, and regional engagement. One can compare institutions to comparable others on a multitude of aspects. In this sense, U-Multirank is perfect for benchmarking. What makes this difficult, though, is that U-Multirank does not provide raw numbers on the different indicators it measures, but divides universities' scores into five groups: 'A (Very good)', 'B (good)', 'C (average)', 'D (Below average)', or 'E (Weak)'. This decision was made because U-Multirank does not wish to create league tables, which its creators consider to be methodologically flawed, and because groupings mean only meaningful differences are shown (Federkeil et al., 2012). This choice indeed avoids false precision (Hicks et al., 2015).

However, it also makes it virtually impossible for an institutional researcher to make a meaningful benchmark. The division into five groups can be arbitrary; it is possible that the difference in indicator score of two universities in group B is greater than the difference between a university in group B and a university in group C. This issue is especially problematic when one does a trend analysis. A clear and meaningful trend is not necessarily translated into a different grouping in U-Multirank, a problem also signalled by Vernon and co-authors (2018). In these cases, it is preferable that raw data are available, so an institutional researcher can make her own custom (trend) analysis with multiple universities and make her own assessment of what differences mean rather than be shown (arbitrary) groups.

Of course, a fair and well-informed benchmark can only be made when scores are good reflections of the indicators they measure. In the following section, I will go even further down the pessimistic road this opinion paper is already walking on, and show how the marketing value of university rankings makes a fair comparison of scores impossible.

Marketing value and gaming

University rankings undoubtedly impact the management and administration of universities. Although many university leaders would prefer to ignore the annual arrival of rankings (and most recognize their fundamental and methodological flaws), they also feel they cannot afford to do so. Or, as indicated by an interviewee in Ellen Hazelkorn's 2011 book on rankings: "it was a case of 'damned if you do, damned if you don't'" (Hazelkorn, 2011, p. 92).

Hazelkorn's book also gives a thorough overview of the effects university rankings have on universities. Many universities have a clear target for their positions in national or

international rankings (Hazelkorn, Loukkola & Zhang, 2014). Universities use their results in rankings for publicity and marketing; a high position helps to attract students and to brand the university to peers and stakeholders (Hazelkorn, 2011). It can also serve as leeway to attract more funding from governments.

Rankings affect strategic, organisational, managerial, and academic actions, such as recruitment and promotional criteria, and the establishment or closing of departments or programmes (Hazelkorn, Loukkola & Zhang, 2014). In some cases, rankings merely acted as an accelerator in these processes – programmes that were already obsolete merely being closed faster (Hazelkorn, 2011). But in other cases, strategic, organisational, managerial, or academic actions were taken with the position on university rankings as the sole purpose. The most famous example is that of King Abdulaziz University in Saudi Arabia, which climbed hundreds of places in the ARWU ranking due to the secondary affiliations it offered to world-leading scientists (Bhattacharjee, 2011).

In addition, universities can nudge their institutional data so they provide an optimal score in university rankings (Hazelkorn, 2011; Sauder & Espeland, 2009). Rankings that rely on self-reported data are a preeminent opportunity and are therefore vulnerable to bias (Vernon, Balas & Momani, 2018). To a lesser extent this also applies to rankings that rely on reputational surveys, as universities can attempt to influence these surveys by marketing themselves to other universities (Espeland & Sauder, 2007) or, in the case of QS, provide contact details of potential respondents (QS, 2018b).

If someone (e.g., an institutional researcher) wants to use the data provided for university rankings, the issue of nudging data is problematic. She can never be sure to what extent a score is ‘real’ – is another university performing better on an indicator or is a score the result of nudging data? The importance of university rankings for publicity and marketing, both to (prospective) students and other stakeholders, makes it unrealistic that these practices will change in the future.

Conclusion and recommendation

I started this paper with the question whether university rankings can be helpful to institutional researchers. The surge of university rankings *has* led to an increased amount of data on research performance – more akin to the standardized data available on teaching. In this regard, rankings could have been helpful.

But ironically, it is the value rankings provide to universities in marketing that prevents rankings from having any analytical value. As institutional researchers, we are involved in a zero-sum game. We have to put in increasing amounts of effort in reporting data to university rankers, because of the publicity and marketing value rankings give to our institutions. However, we cannot reap the benefits of these reporting efforts exactly because of this marketing value; we cannot use the data for our actual work: helping university leaders make informed decisions.

If universities truly wanted to use data for benchmarking, either stricter definitions would be needed (ideally agreed on by universities themselves), or the database would have to be exclusive to users only. An ideal candidate for such a database would be U-Multirank, as its creators already hold a large variety of data and have less of a commercial interest in their current product than the ARWU, Times Higher Education or QS rankers. However, the

question is whether universities would spend resources on such a database. After all, if results are not public, the current main value of rankings for universities disappears.

References

Association for Institutional Research (2018). About AIR. Retrieved March 26, 2018 from: <https://www.airweb.org/AboutUs/Pages/default.aspx>.

Bhattacharjee (2011). Saudi Universities Offer Cash in Exchange for Academic Prestige. *Science*, 334, 1344-145.

Bryant, R. (2017). Institutional researchers and librarians unite! Retrieved March 26, 2018 from: <https://hangingtogether.org/?p=6037>.

Espeland, W.N. & Sauder (2007). Rankings and Reactivity: How Public Measures Recreate Social Worlds. *American Journal of Sociology*, 113, 1-40.

Federkeil, G., Kaiser, F., van Vught, F.A. & Westerheijden, D. (2012). U-Multirank: Background and Design. In F.A. van Vught & F. Ziegele (Eds.), *Multidimensional Ranking: The Design and Development of U-Multirank*, (pp. 85-96). Cham: Springer Publishing.

Haskell, C.D. (2017). Institutional research as a bridge: Aligning institutional internal data needs and external information requirements from a strategic view. *Higher Education Evaluation and Development*, 11, 2-11.

Hazelkorn (2011). *Rankings and the Reshaping of Higher Education: The Battle for World-Class Excellence*. Houndsmills: Palgrave Macmillan.

Hazelkorn, E., Loukkola, T. & Zhang, T. (2014). *Rankings in institutional strategies and processes: Impact or illusion?* Brussels: European University Association.

Hicks, D., Wouters, P., Waltman, L., de Rijcke, S. & Rafols, I. (2015). Bibliometrics: The Leiden Manifesto for research metrics. *Nature*, 520, 429-431.

Leiden Ranking (2017). CWTS Leiden Ranking 2017 – Moving beyond just ranking. Retrieved March 21, 2018 from: <http://leidenranking.com/>.

Levy, G.D. & Ronco, S.L. (2012). How Benchmarking and Higher Education Came Together. *New Directions for Institutional Research*, 156, 5-13.

Leydesdorff, L., Bornmann, L., Comins, J.A. & Milojević, S. (2016). Citations: Indicators of Quality? The Impact Fallacy. *Frontiers in Research Metrics and Analytics*, <https://doi.org/10.3389/frma.2016.00001>.

Marginson, S. & van der Wende, M. (2007). To Rank or To Be Ranked: The Impact of Global Rankings in Higher Education. *Journal of Studies in International Education*, 11, 306-329.

McKenzie, K. & Schweitzer, R. (2001). Who Succeeds at University? Factors predicting academic performance in first year Australian university students. *Higher Education Research & Development*, 20, 21-33.

Moed, H.F. (2017). A critical comparative analysis of five world university rankings. *Scientometrics*, 110, 967-990.

Prewitt-Freilino, P.A. & Rush, N.A. (2017). Data Entrepreneurship: A Necessary Mindset for Institutional Research Effectiveness. *New Directions for Institutional Research*, 173, 11-23.

QS (2018a). University Rankings. Retrieved March 21, 2018 from: <https://www.topuniversities.com/university-rankings>.

QS (2018b). Academic Reputation. Retrieved April 9, 2018 from: <http://www.iu.qs.com/university-rankings/indicator-academic/>.

de Rijcke, S., Wouters, P.F., Rushforth, A.D., Franssen, T.P. & Hammarfelt, B. (2016). Evaluation practices and effects of indicator use—a literature review. *Research Evaluation*, 25, 161-169.

Sauder, M. & Espeland, W.N. (2009). The Discipline of Rankings: Tight Coupling and Organizational Change. *American Sociological Review*, 74, 63-82.

Shin, J.C. & Toutkoushian, R.K. (2011). The Past, Present, and Future of University Rankings. In J.C. Shin, R.K. Toutkoushian & U. Teichler (Eds.), *University Rankings: Theoretical Basis, Methodology and Impacts on Global Higher Education*, pp. 1-16. Cham: Springer Publishing.

Stolker, C. (2014). *Rethinking the law school*. Cambridge: Cambridge University Press.

Times Higher Education (2018). World University Rankings 2018. Retrieved March 21, 2018 from: https://www.timeshighereducation.com/world-university-rankings/2018/world-ranking#!/page/0/length/25/sort_by/rank/sort_order/asc/cols/stats.

U-Multirank (2018). About – The project. Retrieved March 21, 2018 from: <https://www.umultirank.org#!/about/project?trackType=about&signtMode=undefined§ion=undefined>.

Usher, A. (2013). Times Higher Education Reputation Rankings 2013. Retrieved April 9, 2018 from: <http://higherstrategy.com/times-higher-education-reputation-rankings-2013/>.

Vernon, M.M., Balas, E.A. & Momani, S. (2018). Are university rankings useful to improve research? A systematic review. *PLoS ONE*, 13, e0193762.

Wilsdon, J., Allen, L., Belfiore, E., Campbell, P., Curry S., Hill, S., ... Johnson, B. (2015). *The Metric Tide: Report of the Independent Review of the Role of Metrics in Research Assessment and Management*, <https://doi.org/10.13140/RG.2.1.4929.1363>.