

University ranking methodologies. An interview with Ben Sowter about the Quacquarelli Symonds World University Ranking¹

Alberto Baccini*, Antonio Banfi[◇], Giuseppe De Nicolao[□], Paola Galimberti[▽]

Keywords: University Rankings; QS; Ben Sowter

Introduction

University rankings represent a controversial issue in the debate about higher education policy. A growing number of papers is devoted to rankings by focusing on methodological questions about the appropriateness of indicators, weightings, and aggregation methods used (among the others Billaut et al. 2010; Freyer 2014; Jovanovic et al. 2012; Paruolo et al. 2013; van Raan 2005; Safón 2013); others investigated the usefulness and consequences of university rankings (as for example Docampo 2011; O'Connell 2013; Taylor et al. 2013). This growing literature, reviewed by (Hazelkorn 2011; Andrejs Rauhvargers 2011; Andrejs Rauhvargers 2013), witnesses a fierce debate and a lack of consensus not only about methodology but also about policy concerns and unintended effects.

One of the best known university ranking is the Quacquarelli Symonds World University Rankings [hereinafter QS]². QS is a ranking published annually since 2004 by Quacquarelli Symonds ltd, a company founded in 1990 and headquartered in London.

¹ We would like to thank Richard Holmes, who maintains a blog dedicated to university ranking (<http://rankingwatch.blogspot.it/>) for his suggestions for formulating questions.

*

Dipartimento di Economia Politica e Statistica, Università degli Studi di Siena

[◇] Dipartimento di Giurisprudenza, Università degli Studi di Bergamo

[□] Dipartimento di Ingegneria Industriale e dell'Informazione, Università degli Studi di Pavia

[▽] Università degli Studi di Milano

² <http://www.topuniversities.com/qs-world-university-rankings>

QS provides a ranking based on a score calculated by weighting six different indicators: the scores obtained in an academic reputation survey (40%) and in an employer reputation survey (10%), the student to faculty ratio (20%), the citations per faculty according to Scopus data (20%); the international student ratio (5%) and, finally, the international faculty ratio (5%)³.

The 2015 edition, published in October 2015, introduced major methodological innovations and, as a consequence, many universities worldwide underwent major changes of their scores and ranks. In particular, the results raised concerns among Italian stakeholders because almost all Italian universities made a tumble in the rankings with respect to previous years.

Below we publish an interview with Ben Sowter, head of division of intelligence unit of Quacquarelli Symonds, responsible for the operational management of all major QS research projects including the QS World University Rankings⁴. The Interview is mainly focused on QS methodology. We sent questions to Ben Sowter 18th October 2015 and we received the reply 23rd October 2015.

Interview

Q1. This interview is mainly focused on methodology. The QS rankings assign a 50 % weighting to two surveys (Academic reputation, 40% and Employer reputation, 10%). How do you justify the adoption of a such a high weight for these surveys?

Our samples are very large – the survey approach is discipline independent and language independent and has been the backbone of our work since the beginning. The surveys can also be conducted independently from any institution and are free from the challenges associated with self-reported data. Our rankings are inclusive (meaning that we rank all deserving institutions whether they choose to be ranked or not) which means our methodology has to be based as much as possible on data that we can gather from a distance, should we have to. Should other indicators present themselves as robust, appropriate and reliable in the future we may consider reducing the emphasis on the surveys.

Q2. Can you give us information about the sample design, the number of responses and the rate of responses to the various parts of the two surveys, and finally about precautions that you take against universities manipulating the surveys?

A good amount of information can be found on our website on the following two links:
www.iu.qs.com/academic-survey-responses/
www.iu.qs.com/employer-survey-responses/

³ <http://www.topuniversities.com/university-rankings-articles/world-university-rankings/qs-world-university-rankings-methodology>

⁴ <http://www.iu.qs.com/about/meet-the-team/>

Q3. *How does QS check the validity of data about students, international students, and faculty sent from universities? Consider for instance Universidad de Buenos Aires (UBA), where Faculty Student score skyrockets from 17.9 in 2014 to 61.4 in 2015, helping the institution to climb 74 ranks in the overall ranking. Do you think it is possible that both 2014 and 2015 data are valid?*

The data collection process is a massive task. We seek validating sources from national statistics agencies, from ministries, from institution's own websites (to make sure they are telling the world similar things as they are telling us) and anywhere else we can find it. Each year the definitions evolve slightly in order to reduce another aspect of ambiguity that this or that institution bring to our attention. Additionally each year more and more universities engage with us and provide us with more precise data. In the case of some of the larger universities in Latin America, they often have large numbers of non-degree and in some cases even high-school students folded into their published totals which are difficult to disaggregate without the institution's help. In an exercise of this scale, errors can certainly happen, but it's a learning system with more institutions engaged each year. In your final question, the answer is that, in a sense both are valid, but they are not both consistent with our specific definitions. We do our best in this respect, but it's a tough job building the necessary relationships for a shared global understanding and an ongoing journey.

Q4. *Why the raw data are not published to enhance transparency both of ranking calculations and data declared by universities?*

Which raw data in particular? Student, faculty and international numbers are now all published on our website on the individual profiles of each institution. They are also available through our QS World University Rankings App on both iOS and Android.

Q5. *Since there have been several scandals involving the recruiting of international students, academic fraud and abuse of visa status etc (for an example see: <http://www.universityworldnews.com/article.php?story=20120912151836107>) do you consider international students a good criteria of excellence?*

There are always exceptions. International students are a measure of how global a university might feel and how popular an institution is as a destination – international students want to go to good universities and do their own research, so yes, up to a certain point I think that international students, at a macro level, is a relevant way to sort the world's leading institutions from one another – especially given that prospective international students are our primary stated audience. That being said, there are reasons why this measure only carries 5% weight. Our view is that our indicators used together present a more meaningful picture than any of them used apart.

Q6. *The methodology adopted in the latest QS rankings has undergone what you have defined “the single biggest shift in approach since 2007”. In this respect, QS’s warning that “universities which have previously been advantaged – those which are very strong in life sciences and natural sciences – may now rank slightly lower than in previous years” sounds as a deliberate understatement. Indeed, changes were all but slight, as many university changed their rank by tens or hundreds: according to our calculation made on 885 institutions, the average rank change was 37. If we limit the comparison to the first 400 universities, the average rank change is 25, which is still very large. Why QS did not*

include a clear warning that this year's ranking could not be compared with those of previous years? And what do you think about the trustworthiness of so conflicting rankings?

We have avoided side by side comparison and repeatedly stressed to in many of our releases and in engagements with institutions and media that year on year comparison is not valid due to the changes in methodology. In the vast majority of cases the impact of the primary changes has been comparatively slight. Most of the cases where changes are much greater than that are where movements in the citations indicator due to changes have augmented changes in position that were down to other components.

Q7. Among the 2015 changes there is “The extension of our academic and employer survey samples to five years, with the earlier years weighted at 25% and 50% respectively”. This should have a stabilizing effect, yet we observe unexplainable oscillations. For example, for University of Pavia, the 2014 Academic Reputation score is 40.3, while in 2014 it disappears, even if the entire sample (2012-13-14) used in 2014 should enter also the 2015 calculation (with 2012 weighted at 50%). The same happens with the Employer Reputation scores that was 24.1 in 2014 and disappears in 2015. How do you explain these oscillations?

Pavia's survey scores have not disappeared; we just don't publish all the specific indicator scores once we break from specific ranks into ranges (i.e. at rank 400). We had a very strong response in 2015 and adding 2011 back in also makes some difference – which actually means from one year to the next we should expect a little more movement than usual, but the extra years should stabilize the measure moving forward which was the main objective.

Q8. As far as we understand (<http://www.in.qs.com/2015/07/potential-refinements-in-the-qs-world-university-rankings-2015/>), another change is the exclusion of papers with more than 10 affiliated institutions, Can you justify excluding some of the most significant research projects when other ways, such as some form of fractional counting, are available?

We have received a lot of feedback on this development and will take another look and take on board some more expert input. We thought about fractional counting, but rejected it. Consider a “regular” paper with a handful of affiliations. We know that internationally collaborative papers receive more citations than those conducted alone at a single institution. Fractional counting would largely eliminate the citation benefit of collaborating in the first place, and QS is strongly in favour of collaboration. It may be that using fractional counting beyond a certain threshold would be an option we would consider.

However, for the time being consider a paper with 200 affiliations which attracts 1000 citations – our previous model would have attributed all 1000 citations to all 200 institutions. Fractional counting would attribute just 5 citations to each institution. Our current model attributes zero. Whilst there is no argument as to the significance of much of the excluded research, the significance to each contributing institution, were it included on a fractional counting basis would be limited. That said, we will review again and I may have further news for you in advance of next year's results.

Q9. Have you considered doing anything about the growing trend towards multiple affiliation and the buying of secondary affiliations?

The fact that our principal methodology uses a “per faculty” ratio as opposed to a “per paper” approach goes some way to reduce the impact of this kind of tactic, and the fact that citations only contribute to 20% of the methodology limits the overall impact. Multiple affiliations can often be valid, but the buying of secondary affiliations is an issue. One idea might be to include a negative reputation question in our surveys – in principle respondents can choose not to vote for an institution if they consider its reputation a matter of concern, but we don’t provide an outlet for respondents to express concern, as opposed to not expressing support and there is a difference. Given the news coverage around some of these behaviours, the institutions in question might be doing themselves reputational damage and our survey could pick that up.

Q10. *QS claims that it “made adjustments to account for language bias emerging from placing greater emphasis on subjects in which a higher proportion of material is published in language other than English”. In the 2015 ranking there is evidence of a “country bias” in the Citations per Faculty indicator, with China, Australia and UK improving their citation per faculty score, and German, Japan and Italy worsening their score with respect to the 2014 score. How do you explain this? Do you intend to accommodate this country bias?*

You can’t infer a country bias from the new results without being certain there wasn’t one in the previous edition. It is certainly correct that there are instances where a country has, broadly, moved in one way or the other, but in most cases there are exceptions. In general, these shifts are down to the proportion of research in Life Sciences, China has less than 17% of research in Life Sciences which is less than half the global average, whereas Germany has 40% and Japan has 34%.

Q11. *If we take regional distribution of QS Stars as a reasonable proxy of your consulting activity, we find that in 2014, no QS Stars were awarded in Germany, Japan and Italy. Conversely, Australia and UK are the countries with the highest density of “QS Stars”, a certification you have to pay for. QS has been accused of a bias toward the UK and other English speaking countries. If the “methodology shift” had pushed forward countries where most of your revenues come from, could this cause embarrassment to QS?*

There is no connection between our commercial activity and our rankings work and there is no obvious correlation between the impact of these latest changes and the distribution of our clients. We have clients in Russia – which was largely unaffected by the recent changes, there are clients in Canada which generally went down, and very little in China which went up very strongly. Had there been a stronger connection, given the extensiveness of our consultations, we would have still needed to proceed.

Q12. *In order to account for the language bias, you introduced a “sliding scale weight adjustment” in Arts & Humanities and Social Sciences & Management. This adjustment is rather unusual. Rather than inflating citation counts in Arts&Humanities and SocSc&Management for countries disadvantaged by the language bias, it deflates them, while inflating their citation counts in the other three faculty areas (Eng&Technol., NatSc., LifeSc&Medicine). As a consequence, all the citation weights for the five areas are affected. For example, according to our calculation based on your technical explanation, a citation in engineering & Technology in China is weighted 1,44 against 0,90 in South Africa; and conversely a citation in the social sciences and humanities in China is weighted 0,30 against 2,57 in South Africa. Do you think that this normalization is correct? Why did you publish only a very*

scanty technical document instead of a complete explanation of the method and of the rationale on which the adjustments are based?

There are obviously many different ways to run the data. We have settled on this one and run it past our international advisory board, so obviously we think it's good. You are the first and only person who has expressed our explanation to be "scanty". The probability of a citation is directly affected by the volume of publication in the area (in English) in that country. The reality is that just because Scopus doesn't cover it doesn't mean Chinese and German universities aren't producing output in the Arts & Humanities, but that it's not in English and not in Scopus – so there's a coverage issue in those countries in those disciplines. At a global level the default weight for an Art & Humanities citation is 18 – to boost that weight further in countries with low coverage would emphasise potential anomalies and create strange results.

Q13. *Since the adoption of the "sliding scale weight adjustment", in fact every country plays the ranking game with its own rules. Do you think is it still possible to compare institutions of different countries? Are you planning to remove the "sliding scale weight adjustment" from next year's rankings?*

That is one interpretation I suppose, but it's not one that we would subscribe to. Citations are not a perfect or complete measure, and even in a flat foot race, countries are playing "by their own rules" as established by culture, incentives, language, proximity to other prodigious producers and much more. A citation does not represent a constant level of excellence or notoriety independently of discipline or location. Our objective with these latest changes is to make our index as fair and reflective of reality worldwide as possible. It may never be perfect, but we believe this represents a substantial improvement. We do not currently have plans to remove this adjustment.

Q14. *In your latest ranking you put Nanyang Technological University (NTU) ahead of Yale and Columbia. Do you really believe this? If the weight given to a single citation for Singapore is close to the Chinese one (1,44 against 1,16 for US), is NTU's exploit another side effect of the "sliding scale weight adjustment"?*

Citations in different faculty areas have different weights, so I'm not sure which one you are referring to here. Singapore and China bear little resemblance to one another. NTU is ahead of Yale according to our methodology. Is NTU a better university than Yale? I suspect that's more in the eye of the beholder than it might seem – it's certainly better at attracting international faculty and students and better at collaborating internationally. Naturally one is based in the US and serves a larger domestic audience compared to the other which is based in Singapore. Whomever consults a ranking, should always contextualize the results. One of the downsides of running a ranking according to a given, published methodology, is that we have to accept the results it eventually yields, and cannot shape it based on history, tradition, intuition or instinct.

Q15. *How would you respond to critics who claim that the QS stars system and the provision of consultancy services represents an unethical conflict of interest for a ranking business?*

I am not aware of any ranking which is not vulnerable to a conflict of interest argument. A ranking run by a university might be inclined to augment their own position. A ranking run by a media group might be inclined to keep the results fluid to engineer news coverage, or to favour its advertisers. A ranking run by a state organisation may have an overt or covert mandate to favour its

own institutions. The care, attention and ethical standards which any organisation applies to its work is probably proportionate to the risk to the organisation of reputation damage caused by departure from those standards. Of all the organisations involved in university rankings, QS has the most to lose by stepping over that line, and therefore I believe that we are the most resolute in maintaining ethical standards when it comes to the execution of our rankings. Those of us with an active role in compiling rankings have contract clauses that initiate termination if we interfere with their outcomes.

We are a group of passionate and committed individuals who operate with integrity and a strong sense of responsibility.

Q16. *How many universities in Italy have a consultancy contract with QS? Could you disclose the list of these Italian universities?*

We can't provide the specifics of work but the following institutions have been listed in public materials as clients of ours: Istituto Clinico Humanitas, Politecnico de Milano, Università Ca' Foscari Venezia, Università Cattolica del Sacro Cuore and Università di Pisa.

References

- Billaut, Jean-Charles., Bouyssou, Denis, & Vincke, Philippe (2010). 'Should you believe in the Shanghai ranking?' *Scientometrics*, 84(1), 237-263.
- Docampo, Domingo (2011). 'On using the Shanghai ranking to assess the research performance of university systems.' *Scientometrics*, 86(1), 77-92.
- Freyer, Leo (2014). 'Robust rankings: Review of multivariate assessments illustrated by the Shanghai rankings.' *Scientometrics*, 100(2), 391-406.
- Hazelkorn, Ellen (2011). *Rankings and the Reshaping of Higher Education. The Battle for World-class Excellence*. London: Palgrave MacMillan.
- Jovanovic, Milica, Jeremic, Veljko, Savic, Gordana, Bulajic, Milica, Martic, Milan (2012). 'How does the normalization of data affect the ARWU ranking?' *Scientometrics*, 93(2), 319-327.
- O'Connell, Catherine (2013). 'Research discourses surrounding global university rankings: Exploring the relationship with policy and practice recommendations.' *Higher Education*, 65(6), 709-723.
- Paruolo, Paolo, Saisana, Michaela, Saltelli, Andrea (2013). 'Ratings and rankings: Voodoo or science?' *Journal of the Royal Statistical Society. Series A: Statistics in Society*, 176(3), 609-634.
- Rauhvargers, Andrejs (2011). *University Rankings and Their Impact*. Brussels: European University Association.
- Rauhvargers, Andrejs (2013). *Global University Rankings and Their Impact. Report II*. Brussels: European University Association.
- Safón, Vicente (2013). 'What do global university rankings really measure? The search for the X factor and the X entity.' *Scientometrics*, 97(2), 223-244.
- Taylor, Michael, Perakakis, Pandelis, Trachana, Varvara, Gialis, Stelios (2013). 'Rankings are the sorcerer's new apprentice.' *Ethics in Science and Environmental Politics*, 13(2), 73-99.

van Raan, Anthony F. J. (2005). 'Fatal attraction: Conceptual and methodological problems in the ranking of universities by bibliometric methods.' *Scientometrics*, 62, 133-143.