Faculty Area Normalization – Technical Explanation

From 2015, QS has adopted an approach to normalize publication and citation data across faculty areas. There are a number of reasons for this development.

Publication and citation data varies greatly across disciplines. It is possible to run highly complex calculations to adjust for this across large numbers of narrow disciplines – other exercises attempt to balance out between hundreds or even thousands of narrow subject areas. This delivers a degree of precision but is not without pitfalls:

1. Whilst the relative influence of citations in different disciplines can be adjusted for, the "size" of a discipline is typically defined by volume of publications where in reality a large proportion of institutional strength may be in disciplines which don't produce high volumes of journal articles
2. The narrow subject groups can amplify anomalies in some subject areas – particularly those where a large proportion of academic outputs are in forms other than journal articles or with low volumes
3. The definition of subject groups in any attempt at normalisation is somewhat arbitrary – the more groups defined, the larger the number of judgement calls that need to be made. Additionally the smaller the subject groupings, the greater the probability that an article will transcend disciplinary boundaries.

The QS World University Rankings methodology utilizes a Citations per Faculty indicator and thus the objective of this approach is to derive a "Normalized Total Citation Count" (NTCC). The primary approach is to simply equalize the influence of the same five faculty areas that are already deployed in the Academic Reputation analysis:

- Arts & Humanities
- Engineering & Technology
- Life Sciences & Medicine
- Natural Sciences
- Social Sciences & Management

Since 2004, these faculty areas have carried equal weight in our academic survey analysis. The adopted assumption is that, in a typical international comprehensive university, each of these faculty areas represents a roughly equitable share of activity. Looking at the distribution of students might inspire a great emphasis on Arts & Humanities and Social Sciences (data from the Higher Education Statistics Agency in the UK, for example, sees 55% of students studying in these areas) whilst looking at the allocation of research funding would lean towards medicine and sciences where research is, typically, more expensive. Equalizing these faculty areas has always seemed a fair and balanced approach.

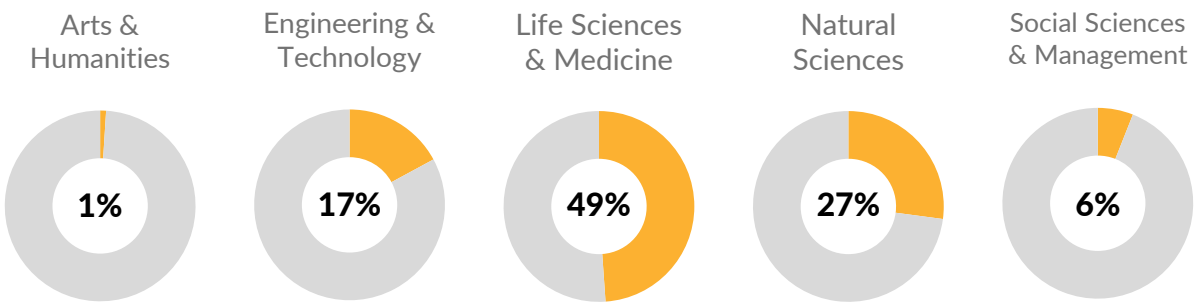


Fig.1 – The distribution of citations across faculty areas in Scopus (2010-2014)

However, it can be seen from Fig. 1 that the distribution of citations across faculty areas is far from equitable, leading to a situation where the Citations per Faculty indicator and, ultimately, the ranking overall favours institutions with a strong emphasis in the sciences.

The new approach described herein adopts the accepted philosophy used in the academic survey analysis and applies it to our citations analysis as well. A key advantage to using these broad sets, as opposed to a narrower discipline focus, is that each will carry greater statistical strength and thus the analysis is less

vulnerable to distortions which have been known to radically affect the fortunes of individual institutions in other analyses based on very small numbers of publications. Indeed, a key reason why QS feels in a position to implement this model from 2015 is due to the growth in the coverage of Arts & Humanities and Social Sciences in the Scopus database over the past few years. The five year window for Arts & Humanities now features over 500,000 citations.

The central intention of this new model, is to equalize the influence of the five faculty areas on the overall outcome of the citations per faculty measure – essentially weighting citations so that each area contributes 20% to the final indicator.

However, given that such a model places greater emphasis on areas where more is published in languages other than English and in forms other than journal articles, we are applying a sliding scale weight adjustment in Arts & Humanities and Social Sciences & Management based on mean productivity levels in those areas for the country where the institution is based.

$$NTCC \equiv \frac{n}{n_{fa}} \sum_{f=1}^{5} C_f w_f a_f$$

n = total citation count prior to normalization

n_{fa} = sum of total citation count across the five faculty areas (typically greater than n since some articles are classified in more than one faculty area)

C_f = count of citations for the given faculty area for the subject institution

w_f = weighting factor for the given faculty area

a_f = weighting adjustment for given faculty area

f = current faculty area, which can be one of ah = Arts & Humanities; et = Engineering & Technology; ls = Life Sciences & Medicine; ns = Natural Sciences; ss = Social Sciences & Management

$$w_f \equiv \frac{n_{fa}}{5x_f}$$

x_f = global count of citations for the given faculty area

$$a_{ah,ss} \equiv \frac{r_f - r_{fmin}}{1 - r_{fmin}}$$

r_f = ratio of a country's papers in the faculty area to the most productive country in the faculty area, in relative terms

$r_{f\,min}$ = the lowest value of r_f across all countries

$$r_{ah,ss} \equiv min\left\{\frac{p_f}{p_{f\,max}} \middle| 1\right\}^{\dagger}$$

$$a_{et,ls,ns} \equiv \frac{5 - (a_{ah} + a_{ss})}{3}$$

p_f = mean proportion of papers from the faculty area for the institution's home country (e.g. in the US 3.69% of papers are attributable to Arts & Humanities and 12.14% to Social Sciences; by contrast, in China, 0.52% are attributable to Arts & Humanities and 4.45% to Social Sciences)

---

† This weighting needs to accommodate specialised institutions in the faculty area to avoid the circumstance where a specialist institution in Economics, in a country where the broader community has lower Scopus counts for Social Sciences, is not disadvantaged. For example, output in Russia is low in Arts & Humanities and Social Sciences and as a result the adjustments in those areas for a typical Russian university will be substantial, but MGIMO, which is an institution focused in those areas will be exempt from the sliding scale adjustment.

$p_{f\ max}$= the maximum value of p_f where the paper count in that faculty area for the given country exceeds the global average (e.g. for Arts & Humanities in 2015 this is South Africa, where p_f = 6.04% and where the paper count is in the region of 5,400 against a global average of 3,400; for Social Sciences $p_{f\ max}$ is also South Africa where p_f = 21.58.76%)

EXAMPLE

This is a worked example based on a well-known UK university using real 2015 data. The institution in question is comprehensive with recognised strengths in Social Sciences and Humanities. The institution accrued 54,586 citations prior to normalization.

n = **44,900,880**

n_{fa} = **52,497,999**

	ah	et	ls	ns	ss
C_f	1,905	12,515	26,764	20,263	8,069
x_f	581,786	9,959,717	25,156,050	13,309,271	3,491,175
w_f	18.05	1.05	0.42	0.79	3.01
r_f	96.0%				81.0%
$r_{f\ min}$	3.0%				10.0%
a_f	95.9%	108.4%	108.4%	108.4%	78.9%
p_f	5.80%	20.25%	35.92%	20.57%	17.46%
$p_{f\ max}$	6.04%				21.58%
$C_f w_f a_f$	32,962	14,303	12,110	17,330	19,144

$$NTCC \equiv \frac{n}{n_{fa}} \sum_{f=1}^{5} C_f w_f a_f$$

$$NTCC \equiv 0.86\ (32{,}962 + 14{,}303 + 12{,}110 + 17{,}330 + 19{,}144)$$

$$NTCC \equiv 81{,}979$$

OVERALL EFFECT

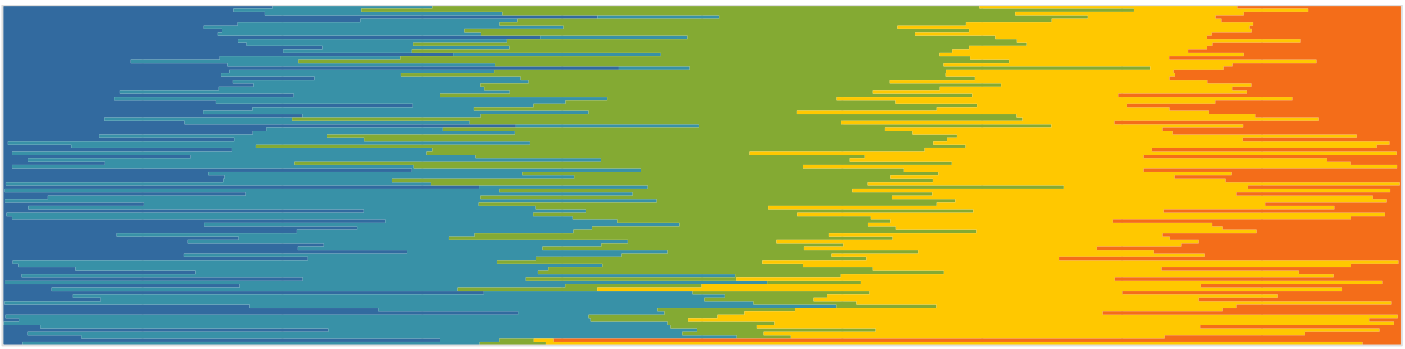
The following charts demonstrate the before and after for the top 100 universities in the overall ranking. The results are sorted based on descending proportion of citations in the life sciences & medicine faculty area

ARTS & HUMANITIES	ENGINEERING & TECHNOLOGY	LIFE SCIENCES & MEDICINE	NATURAL SCIENCES	SOCIAL SCIENCES & MANAGEMENT

BEFORE

1% 17% 51% 24% 6%

AFTER



15% 23% 26% 24% 11%

NOTE:

For clarity, QS bibliometric analysis excludes self-citations and, from 2015, excludes publications carrying more than ten institutional affiliations (at time of introduction this represents approximately 0.34% of publications in Scopus). Affiliated hospitals are included.

Furthermore, after consultation with Elsevier, some content types have been excluded from our analysis this year:

IN	OUT
Article	Abstract Report
Review	Conference Review
Conference Paper	Editorial
Book	Erratum
Book Chapter	Letter
Article in Press	Note
Business Article	Press Release
	Short Survey