

## NORMALIZED CITATION IMPACT SCORE CALCULATION FOR INSTITUTIONAL PROFILES

### Explanation of variance between subjects

The Normalized Citation Impact is calculated from data found in the Thomson Reuters service known as *InCites*<sup>TM</sup> which is in turn derived from the *Web of Science*<sup>SM</sup>. The specifics of the data that are used are as follows:

Databases used = Science Citation Index expanded (SCIE), Social Sciences Citation Index (SSCI) and the Arts & Humanities Citation Index (AHCI)

Document types used = Articles, Reviews and Conference Proceedings (note that we only use conference proceedings from the above databases, we do not use the database known as the Conference Proceedings Citation Index at this time)

The time period is for articles published in the 5 year period up to and including the year that is displayed. For example Normalized Citation Impact for 2010 will be based on publications from 2006-2010. However, it includes citing articles up to the most current year available.

Work is done prior to the analysis to include all publications authored by each university, including those papers where the institutional affiliation uses alternative names, affiliated institutions, old names and other variants.

The data used in Institutional Profiles is identical to the data used in the Global Comparisons module of Thomson Reuters InCites. Additionally, the methodology of calculating the Normalized Citation Impact indicator is identical to the Aggregate Performance Indicator in InCites with the exception of the time period used.

### Normalized Citation Impact (NCI)

The Normalized Citation Impact measures the quotient of an observed citation rate or impact of an institution and an expected citation rate for the institution or country. i.e. it compares the performance of an institution to the average performance of the world .

There are differences in citation rates for research fields and time periods. Therefore, the quotient of the observed citation rate and the expected citation rate must be calculated separately for each field and edition year. If we name the quotients the Relative Citation Rate (RCR) then:

$$RCR_{ift} = \frac{c_{ift}}{p_{ift}} / \frac{c_{ft}}{p_{ft}}$$

Where “c” is the number of citations and “p” is the number of papers. “i” is the data limited to the papers authored by that institution, “f” is the data for a particular field and “t” is the data for a particular year.

In order to obtain the Normalized Citation Impact the weighted mean (weighted with the corresponding number of publications of the institution) is calculated:

$$NCI_i = \frac{\sum_f \sum_t p_{ift} RCR_{ift}}{\sum_f \sum_t p_{ift}}$$

If we denote the expected citation rate by:

$$e_{ft} = \frac{c_{ft}}{p_{ft}}$$

One may rewrite this as:

$$NCI_i = \frac{\sum_f \sum_t p_{ift} RCR_{ift}}{\sum_f \sum_t p_{ift}} = \frac{\sum_f \sum_t p_{ift} \frac{c_{ift}}{p_{ift}} \frac{1}{e_{ft}}}{\sum_f \sum_t p_{ift}} = \frac{\sum_f \sum_t \frac{c_{ift}}{e_{ft}}}{\sum_f \sum_t p_{ift}}$$

### Regional Modification

Once, the Normalized Citation Impact has been calculated we modify the results to take into account the country/region where the institution is based. This is known as a *Regional Modification*.

The concept of the Regional Modification is to overcome the differences between publication and citation behaviour between different countries and regions. For example some regions will have English as their primary language and all the publications will be in English, this will give them an advantage over a region that publishes some of its papers in other languages (because non-English publications will have a limited audience of readers and therefore a limited ability to be cited). There are also factors to consider such as the size of the research network in that region, the ability of its researchers and academics to network at conferences etc and the local research, evaluation and funding policies that may influence publishing practice. For example, in some countries it is policy that doctoral student candidates must publish one or several articles in reputable journals before their doctoral degree can be awarded. This enforced behaviour may have an influence on the citation impact of that country/region; however the exact nature of the influence is unclear.

The Regional Modification takes the form of a moderate modification that in part recognises that the citation impact for some regions will be influenced by the factors discussed above and in part the citation impact will be a true reflection of the research performance in that region. The modification factor is the Normalized Citation Impact for the entire region/country (region in this context is used to identify entities such as the Hong Kong Special Administration Region which has a very different citation trend to mainland China).

The Regional Modification takes the form of:

$$\text{Regional Modified } NCI_i = \frac{NCI_i}{\sqrt{NCI_{\text{country/region}}}}$$

Scores

Once the Regionally Modified Normalized Citation Impact has been generated, this is used to generate scores. Scores are based on Cumulative Probability and are intended to make comparability between different data types. Because the scores are based on a “z-score” type of methodology the scale and distributions of the data sets are irrelevant, which enables comparisons between different indicators.

One important factor to bear in mind is that the scores do not reflect actual performance; instead they reflect the position of one particular university within the distribution of all universities. Therefore it is quite possible that a university that has seen an increase in its citation impact (or any other indicator) may see a reduction in the score if other universities have seen a larger increase or additional universities have been added to the dataset. Additionally it is not true that a score that is double another will have twice the citation impact of the other.

Furthermore, the score is based on the distribution of values for all the universities in the Institutional Profiles project. Currently there are more than 650 universities in the Institutional Profiles dataset and this represents a top slice of some of the highest performing institutions in the world. Scores reflect the performance within that group, not the performance compared to the entire world. So for example, a university may have a Normalized Citation Impact that is above world average (based on a comparison of all the papers in the world) but the score may be low because it is based on a comparison to a set of the highest performing universities.

### Subject Score Calculation

We use the same methodology to assign the values and scores in each of the six subject areas. The first step is to generate the normalized citation impact for the university and then modify it using the regional modification based on the Normalized Citation Impact for the same subject and country/region in which the university is located.

This can result in some surprising results, for example if a university is based in a country that traditionally does not have a strength in a particular subject area they may find that the regional modification gives a larger boost to a particular subject area than one observed for the institution overall.

Next we generate the score based on the distribution of universities in that subject area. This can also generate surprising results as we sometimes see significantly different average values and distributions of values in different subjects. So for example two subjects might have the same value for the normalized citation impact, but very different scores because the distribution of values for all the universities in that subject is very different.

Indeed there are cases where some universities get a higher “Score” for all subjects than they do for the overall, but this is simply a reflection of the fact that the distributions of each of the subjects are different. This is particularly true of Arts & Humanities and Social Sciences where the data is more widely distributed.

### Size Related Issues

There are occasions where a university may see a high citation impact score for a particular subject that is based on a very small number of papers. For example, an Institute of Technology with no Arts & Humanities faculty may actually have a high citation impact for Arts & Humanities because of the crossovers between scientific disciplines. For example engineers publishing in architecture or design journals. This may result in a citation impact (or citations per paper) being generated on a very small number of papers and have a high value.

However, the calculation of the citation impact for the institution as a whole takes into account the volumes in each subject and the influence of a particular subject based on a small number of papers, when combined with other subjects with a larger number of papers is likely to be negligible.