

Ten principles for the responsible use of university rankings

Ludo Waltman, Paul Wouters, and Nees Jan van Eck

Centre for Science and Technology Studies (CWTS), Leiden University

University rankings are controversial. There is a lot of criticism on well-known rankings such as the Academic Ranking of World Universities (ARWU), commonly referred to as the Shanghai Ranking, and the World University Rankings of Times Higher Education (THE) and QS. Nevertheless, universities often feel that they are under pressure to show a good performance in these rankings and universities may therefore pay considerable attention to these rankings in their decision making.

Today the 2017 edition of our CWTS Leiden Ranking was released. We use this opportunity to make a statement on appropriate and inappropriate ways of using university rankings. Below we present ten principles that are intended to guide universities, students, governments, and other stakeholders in the responsible use of university rankings. The ten principles relate to the design of university rankings and their interpretation and use. In discussing these principles, we pay special attention to the Leiden Ranking and the way in which it differs from the ARWU, THE, and QS rankings.

Proposals of principles that could guide the responsible use of scientometric tools are not new. In 2015, we contributed to the Leiden Manifesto, in which ten principles for the proper use of scientometric indicators in research evaluations were presented. The Leiden Manifesto considers scientometric indicators in general, while the principles presented below focus specifically on university rankings. Some of the principles proposed below can be seen as elaborations of ideas from the Leiden Manifesto in the context of university rankings.

Many complex questions can be raised in discussions on university rankings. The ten principles that we present certainly do not answer all questions. The principles are based on our experience with the Leiden Ranking, our participation in discussions on university rankings, and also on scientific literature and policy reports in which university rankings are discussed. We are well aware that the principles may need to be refined and amended. Critical feedback on the principles is therefore very much appreciated, including suggestions on alternative principles.

Design of university rankings

1. *A generic concept of university performance should not be used*

The THE ranking claims to “provide the definitive list of the world’s best universities”. Similar claims are sometimes made by other major university rankings. This is highly problematic. Different users of university rankings are interested in different dimensions of university performance, and therefore a shared notion of ‘best university’ does not exist. Whether a university is doing well or not depends on the dimension of university performance that one is interested in. Some universities for instance may be doing well in teaching, while others may be doing well in research. There is no sensible way in which a good performance in one dimension can be weighed against a less satisfactory performance in another dimension.

The problematic nature of a generic concept of university performance is also visible in the composite indicators that are used in university rankings such as ARWU, THE, and QS. These composite indicators combine different dimensions of university performance in a rather arbitrary way. The fundamental problem of these indicators is the poorly defined concept of university performance on which they are based.

The Leiden Ranking considers only the scientific performance of universities and does not take into account other dimensions of university performance, such as teaching performance. More specifically, based on the publications of a university in international scientific journals, the Leiden Ranking focuses on the scientific impact of a university and on the participation of a university in scientific collaborations. Different aspects of the scientific performance of universities are quantified separately from each other in the Leiden Ranking. No composite indicators are constructed.

2. *A clear distinction should be made between size-dependent and size-independent indicators of university performance*

Size-dependent indicators focus on the overall performance of a university. Size-independent indicators focus on the performance of a university relative to its size or relative to the amount of resources it has available. Size-dependent indicators can be used to identify universities that make a large overall contribution to science or education. Size-independent indicators can be used to identify universities that make a large contribution relative to their size. Size-dependent and size-independent indicators serve different purposes. Combining them in a composite indicator, as is done for instance in the ARWU ranking, therefore makes no sense. In the Leiden Ranking, size-dependent and size-independent indicators are clearly distinguished from each other.

Users of university rankings should be aware that constructing proper size-independent indicators is highly challenging. These indicators require accurate data on the size of a university, for instance internationally standardized data on a university’s number of researchers or its amount of research funding. This data is very difficult to obtain. In the Leiden Ranking, no such data is used. Instead, size-independent indicators are constructed by using the number of publications of a university as a surrogate measure of university size.

3. *Universities should be defined in a consistent way*

In order to make sure that universities can be properly compared, they should be defined as much as possible in a consistent way. When a university ranking relies on multiple data sources (bibliometric databases, questionnaires, statistics provided by universities themselves, etc.), the definition of a university should be consistent between the different data sources. However, even when relying on a single data source only, achieving consistency is a major challenge. For instance, when working with a

bibliometric data source, a major difficulty is the consistent treatment of hospitals associated with universities. There is a large worldwide variation in the way in which hospitals are associated with universities, and there can be significant discrepancies between the official relation of a hospital with a university and the local perception of this relation. Perfect consistency at an international level cannot be achieved, but as much as possible a university ranking should make sure that universities are defined in a consistent way. Rankings should also explain the approach they take to define universities. The Leiden Ranking offers such an explanation. Unfortunately, major university rankings such as ARWU, THE, and QS do not make clear how they define universities.

4. *University rankings should be sufficiently transparent*

Proper use of a university ranking requires at least a basic level of understanding of the design of the ranking. University rankings therefore need to be sufficiently transparent. They need to explain their methodology in sufficient detail. University rankings such as ARWU, THE, and QS offer a methodological explanation, but the explanation is quite general. The Leiden Ranking provides a significantly more detailed methodological explanation. Ideally, a university ranking should be transparent in a more far-reaching sense by making available the data underlying the ranking. This for instance could enable users of a ranking to see not only how many highly cited publications a university has produced, but also which of its publications are highly cited. Or it could enable users to see not only the number of publications of a university that have been cited in patents, but also the specific patents in which the citations have been made. Most university rankings, including the Leiden Ranking, do not reach this level of transparency, both because of the proprietary nature of some of the underlying data and because of commercial interests of ranking producers.

Interpretation of university rankings

5. *Comparisons between universities should be made keeping in mind the differences between universities*

Each university is unique in its own way. Universities have different missions and each university has a unique institutional context. Such differences between universities are reflected in university rankings and should be taken into account in the interpretation of these rankings. A university in the Netherlands for instance can be expected to be more internationally oriented than a university in the US. Likewise, a university focusing on engineering research can be expected to have stronger ties with industry than a university active mainly in the social sciences.

To some extent, university rankings correct for differences between universities in their disciplinary focus. So-called field-normalized indicators are used for this purpose, but these indicators are used only for specific aspects of university performance, for instance for quantifying scientific impact based on citation statistics. For other aspects of university performance, no correction is made for the disciplinary profile of a university. The collaboration indicators in the Leiden Ranking for instance do not correct for this. In the interpretation of the indicators provided in a university ranking, one should carefully consider whether the disciplinary profile of a university has been corrected for or not.

6. *Uncertainty in university rankings should be acknowledged*

University rankings can be considered to be subject to various types of uncertainty. First, the indicators used in a university ranking typically do not exactly represent the concept that one is interested in. For instance, citation statistics provide insight into the scientific impact of the research of a university, but

they reflect this impact only in an approximate way. Second, a university ranking may have been influenced by inaccuracies in the underlying data or by (seemingly unimportant) technical choices in the calculation of indicators. Third, there may be uncertainty in a university ranking because the performance of a university during a certain time period may have been influenced by coincidental events and may therefore not be fully representative of the performance of the university in a more general sense. It is important to be aware of the various types of uncertainty in university rankings. To some extent it may be possible to quantify uncertainty in university rankings (e.g., using stability intervals in the Leiden Ranking), but to a large extent one needs to make an intuitive assessment of this uncertainty. In practice, this means that it is best not to pay attention to small performance differences between universities. Likewise, minor fluctuations in the performance of a university over time can best be ignored. The focus instead should be on structural patterns emerging from time trends.

7. *An exclusive focus on the ranks of universities in a university ranking should be avoided; the values of the underlying indicators should be taken into account*

The term 'university ranking' is somewhat unfortunate, since it implies a focus on the ranks of universities, which creates the risk of overlooking the values of the underlying indicators. Focusing on the ranks of universities can be misleading because universities with quite similar values for a certain indicator may have very different ranks. For instance, when universities in the Leiden Ranking are ranked based on their proportion of highly cited publications, the university at rank 300 turns out to have just 10% fewer highly cited publications than the university at rank 200. By focusing on the ranks of universities, one university may seem to perform much better than another, while the performance difference may in fact be relatively small.

Users of university rankings should also be aware that the rank of a university may drop when the number of universities included in a university ranking is increased. Such a drop in rank may be incorrectly interpreted as a decline in the performance of the university. The value of the underlying indicator may show that there actually has been no performance decline and that the drop in rank is completely due to the increase in the number of universities included in the ranking.

Use of university rankings

8. *Dimensions of university performance not covered by university rankings should not be overlooked*

University rankings focus on specific dimensions of university performance, typically dimensions that are relatively easy to quantify. The Leiden Ranking for instance has a quite narrow scope focused on specific aspects of the scientific performance of universities. Some other university rankings have a broader scope, with U-Multirank probably being the most comprehensive ranking system. However, there is no university ranking that fully covers all relevant dimensions of university performance. Teaching performance and societal impact are examples of dimensions that are typically not very well covered by university rankings. Within the dimension of scientific performance, scientific impact and collaboration can be captured quite well, but scientific productivity is much more difficult to cover. Dimensions of university performance that are not properly covered by university rankings should not be overlooked. Users of university rankings should be aware that even the most comprehensive rankings offer only a partial perspective on university performance. The information needs of users should always be leading, not the information supply by university rankings.

9. *Performance criteria relevant at the university level should not automatically be assumed to have the same relevance at the department or research group level*

Performance criteria that are relevant at the level of universities as a whole are not necessarily relevant at the level of individual departments or research groups within a university. It may for instance be useful to know how often articles published by a university are cited in the international scientific literature, but for a specific research group within the university, such as a research group in the humanities, this may not be a very useful performance criterion. Similarly, one may want to know how many publications of a university have been co-authored with industrial partners. However, for research groups active in areas with little potential of commercial application, this may not be the most appropriate performance criterion. It may be tempting for a university to mechanically pass on performance criteria from the university level to lower levels within the organization, but this temptation should be resisted. This is especially important when the distribution of resources within a university is partially dependent on key performance indicators, as is often the case.

10. *University rankings should be handled cautiously, but they should not be dismissed as being completely useless*

When used in a responsible manner, university rankings may provide relevant information to universities, researchers, students, research funders, governments, and other stakeholders. They may offer a useful international comparative perspective on the performance of universities. The management of a university may use information obtained from university rankings to support decision making and to make visible the strengths of the university. However, when doing so, the limitations of university rankings and the caveats in their use should be continuously emphasized.