Multidimensional ranking: 
a new transparency tool 
for higher education and research

by
Frans A. van Vught and Don F. Westerheijden
Center for Higher Education Policy Studies (CHEPS),
University of Twente, the Netherlands

This paper sets out to analyse the need for better “transparency tools” which inform university stakeholders about the quality of universities. First, we give an overview of what we understand by the concept of transparency tools and those that are currently available. We then critique current transparency tools’ methodologies, looking in detail at the question of data sources, the risks involved in constructing league tables and the challenges in using composite indicators. Lastly, we argue in favour of developing a new principle for transparency tools: that of multidimensional ranking.
Classement multidimensionnel : un nouvel outil de transparence pour l’enseignement supérieur et la recherche

par
Frans A. van Vught et Don F. Westerheijden
Centre d’étude sur les politiques d’enseignement supérieur (CHEPS),
Université de Twente, Pays-Bas

Cet article tente d’analyser la nécessité de créer de meilleurs « outils de transparence » destinés à informer les étudiants et autres acteurs de l’enseignement supérieur sur la qualité des universités. En premier lieu, nous offrons une vue d’ensemble sur ce que nous entendons par le concept d’outils de transparence et sur les outils déjà disponibles à l’heure actuelle. Ensuite, nous analysons en détail les techniques d’utilisation des outils de transparence existants, notamment la question des sources de données, des risques sous-jacents à l’établissement de systèmes de classement et les défis posés par l’utilisation d’indicateurs composites. Enfin, nous exprimons notre avis favorable au développement d’un nouveau principe d’utilisation d’outils de transparence : celui du classement multidimensionnel.
Introduction: the rise of transparency tools

International discussions on higher education and higher education policies have given rise to a new concept: that of “transparency”. Transparency in this context relates to the need to provide information on higher education institutions’ efforts and performance in their various fields of activity. It is also related to the concept of quality assurance. If the latter is perceived as a set of activities intended to provide proof of quality to higher education institutions’ external stakeholders, then creating transparency entails providing the information which these stakeholders need in order to form judgements and take decisions. Such decisions can range from students choosing between specific educational programmes to public or private agencies awarding research contracts and governments deciding on accountability issues relating to funding. Therefore, transparency instruments are information tools designed to communicate information on higher education institutions’ efforts and performance to external stakeholders.

Transparency tools are urgently needed in higher education and more so than in many societal sectors. In economic terms, higher education is either an “experience good” or a “credence good”. In the case of an experience good, the quality can only be judged after it has been consumed (and by implication, after it has been paid for), unlike search goods, whose quality and price are clear in advance. Credence goods are those whose quality remains unknown even after consumption, for example, a medical consultation or computer repairs (Bonroy and Constantatos, 2008; Dulleck and Kerschbamer, 2006). From this perspective we question whether students can ever really gauge to what extent their university experience has enhanced their knowledge, skills and competencies. Also, to what extent is that enhancement specifically attributable to the institution at which they studied? (We distinguish initial from post-initial higher education in this respect; cf. Westerheijden, 2003.) Whether higher education is seen as an experience or credence good, we may safely assume that students cannot know the quality of their educational experience beforehand. Similar arguments can be advanced for other stakeholders in higher education such as businesses, professions and governments. This implies that information asymmetry exists; the value of transparency tools is in rectifying these asymmetries in order to maximise the social benefits of higher education.
Transparency instruments have become increasingly widespread over the last decade. Classifications, rankings and league tables (three of the best-known transparency tools) have become part of the global higher education scene. It is widely recognised now that, although these transparency tools are sometimes controversial, they are both well-established and increasingly influential, impacting on both university and national decision makers. They reflect growing international competition among universities for talent and resources; at the same time their results reinforce competition. On a positive note, they encourage decision makers to heighten their ambitions and to challenge their higher education institutions to perform even better, especially in the case of research universities that dominate global league tables. However, major concerns remain in relation to the league tables’ methodological underpinnings (Bowden, 2000; Clarke, 2002; Wende, 2008; Dill and Soo, 2005; Gottlieb, 1999; Harvey, 2008; King et al., 2008; Marginson, 2008; Usher and Savino, 2006; Dyke, 2005; Yorke, 1998) and to their policy impact on stratification rather than the diversification of their mission (IHEP, 2009; Marginson, 2008; Vught, 2008). Harvey summarised much of this criticism as pointing to a failure to follow accepted methodological rules in relation to deductive analysis (spanning from concept to indicators) and synthesis (spanning from indicators to a composite ranking). Instead, rankings are made from whatever indicators are measurable and available (Harvey, 2008).

While we agree with the principle of conceptual soundness and the need for the deductive development of indicators, the main problem with Harvey’s summary is that there is not a single theory explaining the link between higher education and stakeholders’ information needs. Higher education institutions have many stakeholders and target groups; these have multiple actions and intentions and sometimes lack clarity when expressing their own information needs. Moreover, “Not all nations or systems share the same values and beliefs about what constitutes ‘quality’ in tertiary institutions, and ranking systems should not be devised to force such comparisons” (International Ranking Expert Group, 2006, principal number 5). Given that there are such different needs, different actors require different information on different subjects. Consequently, multi-dimensional transparency tools tailored to individual needs are one obvious solution to complex, unclear information needs.

In terms of multiple stakeholders, we distinguish between:

- Students at the micro level of teaching.
- The same persons at the meso level, i.e. graduates of whole study programmes.
- Employers (businesses and other organisations) who are “users” of graduates.
● To some extent, the same companies who are users of research and innovation.
● Local, regional and national authorities who are often users of higher education institutions’ “third mission” (i.e. their mission to disseminate knowledge and innovation to their environment). In addition, they usually act as the guardians of the broader interests of society in terms of economic and social development and are often major funders.

At the same time, universities’ teaching and research staff are experts par excellence who have cutting edge subject knowledge and, by using educational methods and research technologies, are in the unique position of being able to “delight the customers” (Juran and Gryna, 1988).

In this multi-stakeholders context, quality “is in the eye of the beholder”; it is a multi-dimensional concept which differs according to particular stakeholder perspectives, and it may be more accurate to talk of a “multitude of qualities” (Brennan et al., 1992; Houston, 2008; Westerheijden, 2007). Transparency tools should reflect this conceptual multiplicity in order to communicate higher education institutions’ quality effectively to their users.

An assessment of current classifications and rankings

The preceding analysis requires some conceptual clarifications or at least working definitions. We regard “transparency tool” as an overarching term, covering all means of providing information on effort and performance in higher education. We make an important distinction within transparency tools between classifications and rankings and, within rankings, between league tables and multi-dimensional approaches.

A classification is a system that allocates objects to groups on the basis of their characteristics. Classifications show horizontal diversity, where differences do not imply ordinal scales of “more”, “bigger” or “better”.

Rankings display vertical diversity in terms of performance by using indicators. Most existing higher education rankings take the form of a league table, a single-dimensional list going from “best” to “worst”, assigning ordinal numbers to the entities which relate only to rank and not scales of difference. Other approaches to ranking use:

● multi-dimensional approaches, which do not try to combine education and research rankings, for example, into a single, composite measure and which are often user-driven because they enable an interactive display of data; and/or
● robust group ratings rather than individual rankings, such as in league tables.
On the basis that these other rankings provide more information for a wide field of stakeholders, we argue that they are better, more sophisticated transparency instruments than league tables. This does not mean that we believe that ranking is inherently good or bad, simply that it is possible to define sets of criteria that make ranking-based approaches better or worse.\(^1\) Although there are many existing rankings and classifications, we summarise the major ones in Table 1.

<table>
<thead>
<tr>
<th>Type</th>
<th>Names</th>
</tr>
</thead>
<tbody>
<tr>
<td>Classifications</td>
<td>Carnegie classification (United States)</td>
</tr>
<tr>
<td></td>
<td>U-Map (Europe)</td>
</tr>
<tr>
<td>Global league tables and rankings</td>
<td>Shanghai Jiao Tong University's (SJTU) Academic Ranking of World Universities (ARWU)</td>
</tr>
<tr>
<td></td>
<td>Times Higher Education (supplement) (THE)</td>
</tr>
<tr>
<td></td>
<td>Leiden Ranking</td>
</tr>
<tr>
<td></td>
<td>Webometrics</td>
</tr>
<tr>
<td></td>
<td>CHE Excellence Ranking (European, not global)</td>
</tr>
<tr>
<td></td>
<td>QS World University Ranking</td>
</tr>
<tr>
<td>National league tables and rankings</td>
<td>US News and World Report (USN&amp;WR; United States)</td>
</tr>
<tr>
<td></td>
<td>Time Good Education Guide (United Kingdom)</td>
</tr>
<tr>
<td></td>
<td>Guardian ranking (United Kingdom)</td>
</tr>
<tr>
<td></td>
<td>Forbes (United States)</td>
</tr>
<tr>
<td></td>
<td>CHE Das Ranking / University Ranking (CHE; Germany)</td>
</tr>
<tr>
<td></td>
<td>Studychoice123 (SK123; the Netherlands)</td>
</tr>
<tr>
<td>Specialised league tables and rankings</td>
<td>Financial Times ranking of business schools and programmes (FT; global)</td>
</tr>
<tr>
<td></td>
<td>BusinessWeek (business schools, United States + global)</td>
</tr>
<tr>
<td></td>
<td>The Economist (business schools; global)</td>
</tr>
</tbody>
</table>

Source: Adapted from CHERPA Network (2010), U-Multirank Interim Progress Report.

**Classifications**

A growing desire to come to grips with the diversity of higher education and research institutions has been paralleled by these systems’ growth and the recognition that key stakeholders can no longer be familiar with all institutions individually. In Europe’s state-funded higher education institution (HEI) systems, bureaucratic categorisations have played an important role in this regard. Some countries make a distinction between institutional types such as academic institutes, universities, polytechnic schools, national research centres and higher education colleges. It has become necessary to develop more sophisticated classification instruments in order to distinguish between quality and performance within these categories or across state boundaries, where there are different categories of institutions and degrees. At the same time, increasingly institutional autonomy has led to
strategic diversity, creating a heterogeneous picture of institutional goals, missions and priorities for institutions within the same category. Two classifications used to date stand out on an international level: the US Carnegie classification and the European U-Map classification tool.

The Carnegie Foundation first published its classification in 1973 as a tool for researchers; this subsequently turned into an authoritative concept across the United States and beyond (McCormick and Zhao, 2005). The Carnegie classification was used widely and turned into something of a league table: to be a “research I university” was prestigious, although this was not the objective of its originators. To counteract this tendency, a major system overhaul in 2005 resulted in a multi-dimensional classification. The new classifications were presented as “different lenses” on the higher education landscape:

They are organized around three fundamental questions: what is taught (Undergraduate and Graduate Instructional Program classifications), who are the students (Enrollment Profile and Undergraduate Profile), and what is the setting (Size and Setting). The original Carnegie Classification framework – now called the Basic classification – has also been substantially revised.²

The European U-Map classification methodology (Vught, 2009; Vught et al., 2010) is currently being populated with HEI data.³ U-Map was originally developed as a user-driven, multidimensional European classification instrument allowing all higher education and research institutions to be classified in function of their efforts in six dimensions:

● research
● innovation
● education profile
● student profile
● internationalisation
● regional outreach

Each dimension is made up of several indicators. Ideally these reflect institutional efforts in different areas, the intention being to phase out performance (output) indicators as effort-based indicators become available. At the heart of the U-Map concept is the creation and analysis of specific “institutional profiles”. This enables users to generate institutional “pictures” showing the various indicators in some or all of the six dimensions (Figure 1). U-Map has also developed access to data through two online tools (a profile finder and a profile viewer) which permits stakeholders to analyse institutional profiles and undertake comparative analysis or institutional strategic profiling. U-Map was developed in close co-operation with the
designers of the most recent Carnegie classification and adapted to European circumstances.

We contend that classifications help map HEIs’ horizontal diversity. They address the ways these institutions vary in terms of the efforts they make in their various fields of activities. In this sense they are descriptive: they present HEIs’ institutional profiles by describing their actual behaviour. Classifications do not indicate how well institutions perform with respect to their chosen profile as they focus essentially on institutional effort rather than performance.

**Rankings**

Higher education institution rankings began in the United States in the early 20th century but their exponential growth dates from the turn of the century (Dill, 2006). Recent overviews of existing ranking systems by Shanghai Jiao Tong University⁴ and the Institute for Higher Education Policy (IHEP⁵) list rankings and league tables in more than 30 countries and across all continents, while several countries (including the United States and the United Kingdom) have produced a number of competing rankings. In this context, it is useful to analyse and classify rankings across the following dimensions:

- primary target groups
- producers: public vs. private and not-for-profit vs. private for-profit
level: institutional vs. field-based

scope: national vs. international

focus: education vs. research

We will discuss these dimensions in the following subsections, and then discuss methodological problems regarding data sources, league tables and composite indicators.

**Primary target groups**

Most national rankings expressly aim to inform (prospective) students and their parents about the quality of universities and programmes in their country. It is a challenge for these rankings to balance simplicity and sophistication. One of the principles of “good ranking” as specified in the Berlin Principles (International Ranking Expert Group, 2006) is that rankings should be geared towards their target group, but in this case there are opposing information needs. Prospective students are among those least informed about higher education, and for them, league table information must be simple and focused on education. At the same time, in order to gain legitimacy within higher education communities, more sophisticated information is needed to avoid accusations of oversimplification and insensitivity to sectoral complexity (Federkeil, 2006). As HEIs also have interests to promote other than just their educational programmes, many league tables seek to include research information to help increase the authoritativeness of their rankings. However, this means that until recently ranking systems tended to give the impression that all available indicators must be relevant for all types of users (King et al., 2008).

A similar problem faces international discipline- or field-based rankings, e.g. rankings of business schools which are published by, inter alia, the Financial Times. These international field-based rankings first appeared in internationally integrated professional fields such as business studies. Recently, other rankings have begun to address specific academic fields separately, e.g. the CHE Excellence Ranking and the field rankings in the Shanghai, Taiwanese and THE league tables.

It is important to recognise that not all students are alike: the US News and World Report (USN&WR) ranking and websites such as Studychoice123.nl (SK123) or CHE mainly target students entering higher education for the first time, i.e. typically adolescents completing secondary education. The Financial Times targets adults with at least several years of professional experience wanting to gain specific skills. These are groups of completely different learners with their own study cost/benefit calculus, with different teaching expectations, different knowledge about HEIs, different information needs and different social preferences and mobility. Given the importance of
responding to diverse student needs, good rankings are those which provide robust information about investment considerations (e.g. future job chances; gaining specific competencies; having well-known teachers [e.g. in performing arts] or additional income [Westerheijden, 2009]). Good rankings should also provide information on consumption considerations (e.g. the availability of sports facilities at HEIs; the possibility of living on campus; acquiring a broad academic education). Considerations relating to consumption may be particularly important to first-time students, while those related to investment may be more important for returning students.

It is important to note that even specific target groups have multiple information needs. In addition, socially-excluded groups’ needs in relation to HEIs are severely constrained by a lack of contextual information such as the kinds of benefits or qualities they might acquire by attending university. League tables further contribute to a stratification of the student body in that it is those with the highest social capital who tend to take them most into account when making strategic higher education choices.

International or global rankings of HEIs do not generally refer explicitly to a defined target group, rather a broad audience closely or loosely involved with higher education. The most prominent global league table, Shanghai’s Academic Ranking of World Universities (ARWU), was initially designed to compare Chinese universities’ research performance in the fields of science and technology, and to compare performance between the Chinese national higher education sector and the rest of the world, especially the United States. The ARWU was primarily a tool for steering national research policy and planning; it therefore targeted policy makers and public authorities (in particular the ministries of education, science and technology).

Detailed research rankings such as the Leiden Ranking (which analyses university performance by using citation data) target university managers (i.e. institutional leaders and their support staff) by giving them information on similar kinds of universities which are either their competitors or against which they might mutually benchmark. Similarly, the Webometrics ranking informs institutional leaders about their higher education or research institutions’ relative prominence on the web, which is potentially useful for informing decisions concerning open access publishing. Both of these focused rankings (the Leiden Ranking and Webometrics) are designed to answer institutional staff’s specific information needs associated with strategic decision making.

Rankings are also of interest to those beyond the primary target group. For example, politicians consult research-based league tables or rankings targeted at prospective students. Leaders of HEIs are generally interested in all types of rankings because of their impact on institutional prestige, and in
particular the negative prestige that results from low rankings. This raises questions about universities’ impartiality as data suppliers, and alerts us to the risk that some institutions might be tempted to “adjust” the rankings in order to enhance themselves or discredit their competitors (Bowman and Bastedo, 2010; Dill and Soo, 2005; Hazelkorn, 2007; Sharp, 1995).

**Producers of rankings**

The majority of national rankings are produced by media companies, a trend started in the United States with the *US News and World Report*. Previously, some rankings had already been produced by academic institutions themselves (e.g. in economics), but in contrast to the media rankings they did not receive substantial public attention. Other examples of media rankings include *The Times* Good University Guide, league tables compiled by the UK newspapers the *Guardian* and *Independent*, national rankings in France (*Nouvel Observateur*) and Italy (*Sole 24 Ore*). As commercial publishers are interested in profit and sales, this raises the risk that they are less concerned by sound and stable methodology than garnering spectacular headlines (Gottlieb, 1999; Stella and Woodhouse, 2006).

A minority of rankings are published by independent, national non-profit organisations. These include the CHE rankings in Germany (although they have a media partner, *Die Zeit*), Studychoice123 in the Netherlands and a Polish ranking published by the Perspektywy Foundation. Furthermore, there are a few examples of national rankings published by public institutions, e.g. the Higher Education Evaluation and Accreditation Council of Taiwan (HEEACT) and the Nigerian Rectors’ Conference.

There are six “global” rankings; these are differentiated in function of their geographical and disciplinary coverage. Three of them are compiled by academic institutions (CWTS/Leiden University, *École des Mines* and Consejo Superior de Investigaciones Científicas [CSIC] in Spain). Whilst ARWU was initially produced by the Shanghai Jiao Tong University (SJTU), the unit responsible for it became independent in 2009 and formed the Shanghai Ranking Consultancy. The *THE* ranking is an outlier among global rankings as it is generated by a major newspaper. Since *THE* changed its data supplier in 2009 in order to have more direct access to citation data and develop its reputation survey, its previous partner (QS) launched its own product, World University Rankings, so as of 2010 there are two commercial global ranking producers.

Chinese Taipei’s HEEACT and the Dutch consortium SK123 (which includes many categories of stakeholders and benefits from a governmental subsidy) seem to be the rankers most closely related to public authorities. They are not involved with policy making in relation to higher education and
research institutions, although some branches of HEEACT work on evaluation and accreditation issues. Most academic institutions which produce global league tables are public actors, but both CHE and Perspektywy are public-private partnerships; they are independent non-profit organisations with close ties to national rectors’ conferences.

Increasingly, authors of rankings are revealing information on their methodology, often on their websites. However, as the AUBR Expert Group put it, it was “found that the results of the Shanghai Jiao Tong Academic Ranking of World Universities (ARWU) are not replicable, thus calling into question the comparability and methodology used” (AUBR Expert Group, 2010, p. 56).

**Institutional and field-based rankings**

In broad terms, it is possible to distinguish between users who are interested in institutional rankings and those interested in “fields”, which we define as smaller organisational units like faculties, schools or departments focusing on a single area of knowledge. Fields may include academic disciplines like economics and physics, interdisciplinary areas like business studies and nano-technology and single study programmes or research programmes in a given area. Field-level rankings are of particular interest to students or individual researchers looking for a study or research base, since programmes across institutions may have quite different qualities. Indicators only showing averages for whole institutions mask particularly strong or weak programmes, implying that for these users institutional rankings are irrelevant or even misleading.

Institutional-level rankings, on the other hand, are popular with government policy makers and institutional leaders who have a legitimate interest in overall characteristics at the institution level. In fact, some characteristics (such as an institution’s mission or its policies relating to participation) only apply at this level. Likewise, the institutional level is useful for mapping more detailed characteristics and identifying similar institutions for co-operation or benchmarking exercises. Besides, policy makers often limit themselves to the institutional level, because it is here that they make policy and funding decisions.

Most global rankings (such as ARWU, THE, QS, Leiden, HEEACT, Webometrics) rank whole higher education and research institutions, and it is in this connection that they are generally associated with the “reputation race” (Vught, 2008). ARWU, HEEACT and THE rankings also publish results for broad fields, but they do not identify the “best institutions in the world” (or whatever designates “best” according to the indicators used) for these fields. Instead, they create a new rank order of institutions from their general ranking in function of individual fields. This means that theoretically the
rankings may miss out the best institutions if these do not figure in the overall top scores. This way they put small, specialised institutions at a disadvantage. Some global rankings only focus on one specific field: the Financial Times ranking is solely concerned with business studies, irrespective of whether these are taught in comprehensive institutions or business schools. In that sense, the FT ranking straddles the field vs. institution divide.

Several national rankings also focus on institutions as a whole, such as USN&WR and Perspektywy. More typically, however, national rankings such as CHE Ranking and SK123 are geared to help prospective students make an informed study programme choice in step with their individual preferences.

National vs. international rankings

The earliest rankings which were published compared colleges across 50 US states. This was comparable to a national level from a technical point of view, but in fact 50 states and around 4 000 HEIs constitute as large and as complex a higher education system as the European Higher Education Area, which comprises 47 countries and several thousand HEIs. Nevertheless, the USN&WR league tables are more like national rankings in other countries and aim to inform US students about the best study options available across the whole country. Similarly, at field level, the CHE Ranking aims to inform students looking for undergraduate study programmes across all 16 German federal states. These examples point to the fact that national-level rankings tend to be designed for a clearer purpose and with a more focused target group in mind than global ones.

International rankings like the ARWU and THE primarily rank whole higher education institutions. They are currently the most controversial and most discussed rankings, as they are said to make or break institutional reputations. Other international rankings such as the Leiden Ranking or Webometrics are more explicit about the limitations of their scope and claims, i.e. research performance and impact (Leiden Ranking) and web presence and activity (Webometrics). That way, they evoke less vehement debate than the former two.

There is a growing demand for more international transparency in the context of international mobility of students, and here we see two primary trends. First, national rankings are expanding to neighbouring areas. Thus, the CHE ranking now includes institutions from inter alia Austria, Switzerland and the Netherlands, although a pan-European ranking does not exist yet. Second, more focused international rankings are beginning to emerge: the CHE Excellence Ranking, for example, focuses on the European market for Masters and PhD students. First piloted in 2007, this covers a limited number of fields
and is restricted to international, research-oriented universities throughout Europe.

**Education vs. research**

Rankings can have two different foci. First, they may set out to provide information about either education or research. For instance, the professed aim of the CHE rankings and SK123 is to inform students about the best places to study, indicating a focus on education. Likewise, the Leiden Ranking explicitly focuses on the research performance of HEIs. Also, the original objective of the ARWU was to inform policy makers in China about the position of Chinese research universities (mainly in the fields of science and technology).

Second, another facet of the word “focus” relates to the **actual indicators** used to compose the ranking. A good ranking system is one in which the indicators used to establish a ranking (*e.g.* research productivity and impact) correspond to the focus the ranking claims to have (*e.g.* to inform institutional leaders about the research strengths of their institution). Most national rankings which focus on informing (prospective) students do indeed have indicators on teaching or use a mix of education and research indicators (plus some context variables). The majority of indicators used in the six main global rankings give the ranking a tendency to evaluate research. Given its objective, this practice makes the Leiden Ranking a good one, but the same practice can become problematic if research-based rankings are used in any way as equivalent measures of general prestige or even quality.

**Methodological issues**

A substantial part of the current discussions on the “use and abuse” of ranking and categorisation relates to a set of methodological issues. In this section we consider three issues that appear to be at the heart of these discussions, namely data sources, league tables and composite indicators in rankings.

**Data sources**

When reviewing the workings of the main ranking methodologies it appears that the data can be broken down into three broad categories: databases and statistics (national and international), data elicited from the institutions to be ranked (“self-reports”) and surveys from different types of respondent.
**Existing databases**

The ideal solution from the point of view of validity, reliability and parsimoniousness of data collection (not bothering higher education institutions with unnecessary questionnaires) is to use an existing database. An instructive example of this is the Carnegie Classification, which distinguishes higher education institutions in function of a large number of criteria and indicators. The information needed to construct these indicators is derived almost uniquely from a publicly available US-wide database, the Integrated Postsecondary Education Data System (IPEDS). IPEDS is based on surveys conducted annually by the United States’ Department of Education’s National Center for Education Statistics (NCES). It is a legal obligation for HEIs benefiting from federal student aid programmes to report enrolments, programme completion, graduation rates, staff, finances, tuition fees and student financial aid data to IPEDS. IPEDS checks the quality of the self-reported data, which means that the Carnegie Foundation is neither burdened with data collection nor its verification. Institutions only have to report data once, rather than to both the government and the Carnegie classification. Almost all data used is freely available at the federal level in the United States.

The system described above is ideal for a classification or ranking, but is rare. Databases exist in a number of countries, but few exist on an international level. The European U-Map project has been disadvantaged in relation to the Carnegie Foundation given the absence of a Europe-wide database on individual higher education and research institutions. U-Map has had to work with national databases and statistics, which raises two difficulties. First, national-based data are notoriously difficult to gather internationally, given the plethora of national sources governed by different data protection laws. They are also hard to compare, due to different data being collected across countries, which moreover often use different definitions of what seem to be the same data. International publications on statistics, such as the OECD’s annual *Education at a Glance*, therefore have an abundance of footnotes in every table or indicator, showing the limits of comparability of international statistics. A final problem that arises is that international databases tend to gather information at the national system level rather than for individual institutions or at the field level within institutions.

One area where it has been easy to effectively compare international data is in the field of bibliometric indicators. The two major databases used for large-scale comparative bibliometric studies are Web of Science (WoS) and Scopus, which cover journal articles published in peer-reviewed journals. Nevertheless, the use of such data comes with its own problems, since publication cultures and modes vary considerably between different fields.
(e.g. Moed, 2005). Peer-reviewed journals are the prime vehicles for knowledge dissemination in the natural sciences, medical sciences and life sciences. However, in many applied sciences and in engineering, conference proceedings are more important than journal articles. In the social sciences and humanities, book publications play an important role. Focusing the data collection on journal articles (which are easily gathered) creates a bias in favour of the sciences and medicine. The corollary of this is that existing indicators penalise universities which are leaders in fields where journals are less important. However, both WoS and Scopus databases are rapidly repairing major lacunae by increasing coverage of journals in under-represented domains of knowledge production, as well as that of international journal coverage and conference proceedings. Nonetheless, the coverage of both databases is likely to remain unsatisfactory for the arts and humanities in the foreseeable future.

A further problem comes from the fact that the journals included in the databases have so far been biased towards English language publications. Output from non-English speaking countries is therefore under-represented, as are publications from major countries which have a long tradition of scientific research (e.g. France, Germany, China and Japan). A further drawback is that this language bias also contains a disciplinary bias. The sciences are mostly international (English) in their modes of publication, while several other fields in the humanities and social sciences have much stronger national research cultures, publishing in the languages of the countries involved. Citation indicators in English therefore reinforce a bias towards sciences. WoS and Scopus have been trying to address this, in particular by developing alternative databases covering humanities journals in more European languages.

Bibliometry also suffers from biases. For instance, publications by large universities are cited more often than those by small ones. Also, the process of cleansing data to assign publications to persons and institutions (or to ascribe web pages to institutions), is far from straightforward, and this lowers the reliability of indicators unless rankers use them with great care (Moed, 2005; Raan, 2005).

**Self-reports from higher education and research institutions**

Where (international) databases fall short, self-reporting by higher education and research institutions becomes the most used data source (Thibaud, 2009). Many types of data are efficiently gathered from institutions themselves such as staff composition, institutional facilities, budget reallocation and licence income, if this is not available elsewhere. Institutions’ virtual monopolies on such data create a “principal-agent” problem and open the door to rankings being “adjusted” by manipulating data. Individual
institutions’ definitions may differ, and almost certainly differ across countries. Normalisation of data to a single, globally-used definition is rarely straightforward. For these reasons, the plausibility of self-reported data needs to be externally tested and validated. Analysis also needs to be carried out on extreme cases e.g. through using time series data, triangulation with other data sources and using the expertise and knowledge of an advisory board.

Surveys

A number of rankings use data resulting from peer surveys (i.e. THE, QS, USN&WR and CHE) notably on institutions’ reputations, as well as information on satisfaction levels from surveys among students and graduates (CHE, SK123).

The degree of satisfaction with which students (and graduates) rate their university experience is a category of data which is particularly valid for rankings that address prospective students (quality “is in the eye of the beholder”). Experience from the CHE and SK123 rankings as well as national student surveys in the United Kingdom and Australia show that student and graduate surveys produce robust comparative information about higher education programmes. But to date little work has been done on the international comparability of this type of survey data. Clearly, it is affected by a range of country-specific factors such as culture (notably the acceptability of critiquing teachers) and scales (Westerheijden et al., 2008).

Another variable for which surveys are often used is institutional reputation. An analysis of CHE data on the reputation of German, Swiss and Austrian universities (Federkeil, 2009) showed that reputation has to be treated with extreme caution in international rankings. Clearly, peers can be biased as a result of unawareness of all but the most famous international higher education and research institutions. The quality of the results depends heavily on the quality and size of the sample. In 2009, a heated discussion even broke out in the United States about the trustworthiness of USN&WR’s peer reports: respondents to the ranking survey were accused of downgrading other higher education institutions in order to improve their own institutional standings.14

League tables

A focus on a specific category of data leads to a very restricted concept of quality, and this narrowness has consequences. Nevertheless, such a narrow concept of quality is what characterises many current rankings, where the focus is on “world class research universities”. Governments have made significant efforts to build “world-class universities” through special funding. They have also been stimulating mergers or taking other measures in favour
of these universities (Salmi, 2009). This approach has been criticised for concentrating efforts at the expense of interest in, and resources for, other parts of higher education systems. Similarly, AACSB, a specialist accreditation organisation in the field of management, criticised MBA programme rankings for taking the risk of narrowing the diversity of business studies to a single, highly specific course, the MBA (AACSB, 2005, p. 7).

Most rankings, both national and international, are published as league tables and to be ranked at the top implies higher quality. This kind of league table is ranked on an ordinal scale, and this poses problems. A change in rank does not necessarily signify a change in institutional quality if the performance of others has changed. Further problems can arise when ranking differences are within the margins of error of the methodologies used. Also, there is a strong risk of random ranking fluctuations arising from measurement errors. For example, in the 2008 THE World Rankings, the difference between the universities ranked number 27 (Brown university) and 43 (University of Queensland) was only 4.5 points, and between numbers 50 and 100 it was only 10 points. In the QS rankings, between 2009 and 2010, 10 institutions varied more than 20 places in the league tables. Hence, these tend to exaggerate differences between institutions and put an extreme emphasis on vertical stratification.

**Composite indicators**

Composite indicators in rankings are scores that are derived from combining a set of underlying variables which produce a single score. Obviously, the choice of variables and the weighting given to individual variables influence the outcomes, as well as the rankings, which emerge in these situations. This raises three issues which need further reflection.

First, assigning weights to individual indicators needs a conceptual model with a set of arguments about the relative weight of the indicators when defining quality. A 1997 study by the National Opinion Research Center on the USN&WR rankings confirmed that “the weights used to combine the various measures into an overall rating lack any defensible empirical or theoretical basis”.\(^\text{15}\) To this day, the problem persists that there are neither generally accepted theoretical nor definite empirical arguments for assigning particular weights to individual indicators (Dill and Soo, 2005). Moreover, the weighting choices create an implicit bias in favour of specific types of higher education institution. Reputation, as measured by international surveys, is for instance enhanced by HEIs being located in a major, well-known city – and by establishing a university brand (Marginson, 2008).

Second, rankings target different groups and these have different priorities and preferences when it comes to comparing universities and
making choices. Moreover, individual users have different priorities and preferences, e.g. prospective students use heterogeneous criteria for selecting a university. To be relevant for users’ decision-making processes, ranking systems must account for this heterogeneity and potentially leave the decision about the relevance – and weighting – of indicators to the users. A composite indicator with fixed weights risks patronising the users of rankings by predetermining the importance and relevance of different indicators. Eccles (Eccles and Gootman, 2002) pointed out that fixed weights also fail to acknowledge non-traditional students' interests, as these may have priorities and interests different from “mainstream” weighting systems. To offer a more user-driven approach, some web-based rankings have introduced an interactive tool in order to let users decide about the relevance of indicators. Some rankings (including the Guardian) do this by allowing the user to assign their own weights to the indicators forming the composite indicator. Others such as SK123, CHE and HEEACT allow users to give priority to a number of indicators, resulting in a personalised ranking of programmes or institutions.

Third, the methodology used by the THE and Shanghai Jiao Tong University rankings to construct their composite indicator has been analysed statistically and was found to be relatively unrobust (Saisana and D’Hombres, 2008). Through a sensitivity analysis and simulations using different weightings, Saisana and D’Hombres showed that 67% of universities in the THE ranking and 60% in the Shanghai Ranking were highly sensitive to the composition of the overall score. The variation of league table positions according to different indicator models in general is greatest in the lower ranks, but even the Massachusetts Institute of Technology drifted from the 10th to the 25th position according to THE data (ibid., p. 53). Saisana and D’Hombres concluded that “no conclusive inference regarding the relative performance for the majority of the universities can be drawn from either ranking” (ibid., p. 8).

Discussion

All over the world higher education and research systems are becoming more complex. Stakeholders are looking for valid and reliable information about their own systems but increasingly also across national borders. More national and international classifications and rankings are being produced to respond to this need for greater transparency.

Asbhys well-known Law of Requisite Variety makes us realise that the more complex higher education systems become, the more complex our way of looking at them needs to be. In other words, for a simple system of uniform universities, perhaps all we need is an elementary league table. However,
given contemporary complexities, more complex instruments are needed if we are to obtain the degree of transparency we are looking for (cf. Vught, 1993). Sophisticated transparency tools are also needed because the role of higher education in society is expanding. As different kinds of stakeholders come into contact with universities they bring their own information needs, which transparency tools need to serve if they are to effectively meet their own goals.

Transparency tools are designed to help stakeholders form judgements and take decisions. In this context, information needs among different stakeholders in higher education are diverse (see, for example, AUBR Expert Group, 2010). Moreover, stakeholders are not homogeneous groups: specific categories of professionals may need tailored study programmes, quite unlike those of secondary-school leavers. Transparency tools therefore must be designed with flexibility to cater for these different needs.

Our analysis builds on a growing literature that critiques existing transparency tools. Our aim, however, is not to suggest that we should stop publishing rankings: many groups of stakeholders show a clear and legitimate need for information on higher education, research institutions and study programmes. Transparency tools such as classifications and rankings therefore have an important role to play in the further development of the external quality assurance of higher education and research. The way forward, in our opinion, is to improve transparency tools, and there are two avenues to achieve this.

The first requires rankers to “self-regulate” by articulating more clearly the principles of ranking and what constitutes “good” ranking. With this in mind, a number of rankers and experts have established the International Ranking Expert Group (IREG), which convenes regularly. At its meeting in Berlin in 2006, it agreed on a set of basic principles for good ranking practice, commonly called the Berlin Principles (International Ranking Expert Group, 2006). The Berlin Principles set out to establish good practice in relation to the purposes and goals of rankings, the design and weight of indicators, the collection and processing of data and the presentation of ranking results. Accordingly, they call for:

- Clarity about purpose and target groups.
- Recognition of the diversity of institutions.
- Transparency in ranking methodologies.
- The measurement of outcomes, not inputs.
- Providing consumers with a clear understanding of how a given ranking was developed and offering them a choice of how rankings are displayed.
- The application of quality assurance principles to the ranking itself.

(N.B. From 2011 onwards, the IREG plans to audit and recognise rankings.)
In general, the Berlin Principles are perceived as relevant guidance on what should constitute “good” rankings. Their application reflects the multidimensional understanding of the concept of quality.

The second avenue is via the recognition of the multiplicity of stakeholders’ interests in higher education and research and hence of the importance of creating multidimensional ranking tools. At CHEPS we are currently preparing a feasibility study to do exactly that, i.e. develop a multidimensional global ranking tool that is able to address different stakeholders’ needs. The study in question is the “U-Multirank” project, which is undertaken by an international research consortium. In addition to CHEPS it involves producers of well-established specialised rankings, namely CHE (which contributes experience with student-focused programme rankings), and CWTS (from the Leiden research ranking). Applied research and innovation – something which has hitherto not featured widely in ranking – is represented in U-Multirank by INCENTEM, which is located at the University of Leuven in Belgium. The Paris-based Observatoire des Sciences et des Techniques (OST) is also part of the research team. The researchers’ diverse areas of expertise reflect the project’s goal, which is to design in a deductive way a ranking that reflects all missions of higher education and research institutions, while recognising that not all institutions need to perform on all dimensions.

A deductive rather than opportunistic approach calls for bringing together existing indicators (that effectively measure what matters) and developing new ones, such as regional impact indicators, where they do not. The project involves a pilot field test to explore the extent to which indicators may be developed in such a way that they reflect higher education and research institutions’ performance in the most varied contexts. The ranking exercise seeks to reflect various stakeholders’ needs, developing both focused rankings for higher education and research institutions as a whole (for institutional decision makers and national policy makers) and field-based rankings to inform (prospective) students.

This is clearly a large, difficult and potentially risky undertaking, and the fact that existing rankings have focused on available data underscores that “measuring what matters” is extremely difficult. To ensure legitimacy as well as validity, the development of the U-Multirank project is embedded in a context in which both international experts and stakeholders are represented. The final report will be published in 2011 and will furnish empirically-grounded answers to the question of whether it is possible to make higher education and research transparent on a world-wide scale.

Only future research will tell whether it is possible to design and implement a multidimensional ranking tool on an international scale. But
such a tool will only make higher education and research transparent if it can also address the various methodological problems discussed above. In addition, this tool should at least address the following three issues. First, higher education institutions should be allowed to present their own specific “institutional performance profiles”. Second, rankings should be able to meet different stakeholders’ needs, and third, higher education and research institutions should be able to showcase their strengths. If these three conditions can be fulfilled, then multidimensional ranking may well mark the next generation of transparency tools.

The authors:
Frans A. van Vught
Honorary Professor and former President of the University of Twente
Center for Higher Education Policy Studies (CHEPS)
University of Twente
P.O. Box 217
7500 AE Enschede
The Netherlands
E-mail: f.a.vanvught@utwente.nl

Don F. Westerheijden
Senior research associate
Center for Higher Education Policy Studies (CHEPS)
University of Twente
P.O. Box 217
7500 AE Enschede
The Netherlands
E-mail: d.f.westerheijden@utwente.nl

The authors wish to thank Dr. Paul Benneworth of the Center for Higher Education Policy Studies (CHEPS) for his contribution in finalising this text. They would also like to mention that parts of this text are drawn from two project reports – to which they contributed – which are available online: Design and Testing the Feasibility of a Multi-dimensional Global University Ranking (CHE et al., n.d.) and U-Multirank: Interim Progress Report (CHERPA Network, 2010).

This project has been funded with support from the European Commission.
Notes

1. This is complicated by the fact that “ranking” may be a noun or a verb, while there is no corresponding verb for “league table”; some confusion stemming from our use of verbs may be unavoidable. More confusion comes from many league tables being called “rankings”.


3. See: www.u-map.eu.


6. For an overview, see www.find-mba.com/mba-rankings.


10. Reputation is an efficient, therefore attractive indicator of “quality” for actors who do not have the time or resources to search for detailed information (Stigler, 1961). Therefore, Rankings enjoy a high level of acceptance among stakeholders and the wider public because of their simplicity and consumer-type information (AUBR Expert Group, 2010).


13. The EU has launched a project to investigate options of getting institutional-level data; see the EUMIDA project (www.eumida.org). Also, the AUBR Expert Group recommends setting up a regular observatory on research information from HE and research institutions (AUBR Expert Group, 2010).


16. See: www.u-multirank.eu. This article reflects the views only of the authors. The Commission cannot be held responsible for any use which may be made of the information contained therein.

References


Brennan, J. et al. (1992), Towards a methodology for comparative quality assessment in European higher education: A pilot study on economics in Germany, the Netherlands and the United Kingdom, CNAA/CHEPS/HIS, London/Enschede/Hannover.


King, R. et al. (2008), Counting what is measured or measuring what counts? League tables and their impact on higher education institutions in England, April 2008/14, Higher Education Funding Council for England, s.l.


Moed, H.F. (2005), Citation Analysis in Research Evaluation, Springer, Dordrecht.


Vught, F.A. (van) et al. (2010), U-Map: The European Classification of Higher Education Institutions, CHEPS, Enschede.


