

UNIVERSITY RANKINGS: A CRITICAL VIEW

Felipe Martínez Rizo*

Translator: Pablo Contreras Fresán
E-mail: deepcolearning@gmail.com

REVISTA DE LA EDUCACIÓN SUPERIOR
ISSN: 0185-2760.
Vol. XL (1), No. 157,
Enero - Marzo de 2011, pp. 77-97

* Universidad Autónoma de Aguascalientes. Education
Department.
E-mail: felipemartinez.rizo@gmail.com

Approach: 05/11/10 • **Acceptance:** 14/12/10

Resumen

Después de repasar la historia de los *rankings* de universidades y describir la forma en que han proliferado, el trabajo presenta las características de tres de los más importantes en el plano internacional: el de Shanghai, el del *Times Higher Education Supplement* y el de las universidades en la *web*, que se contrastan con el sistema del *Center for Higher Education Development* alemán. Se presentan los criterios que tales ordenamientos deberían satisfacer que para ser considerados consistentes, se aplican a los ordenamientos en general y a dos de los mencionados en particular, y se concluye que sus limitaciones metodológicas son tan graves que no justifican su pretensión de ser confiables para evaluar a las universidades.

Palabras clave:

- Educación superior
- Universidades
- Evaluación

Abstract

After reviewing the history of university rankings and describing their growth, the characterization of three of the most important rankings internationally is presented, namely: Shanghai, *Times Higher Education Supplement* and the *Web of World Universities*, these are contrasted with the German Center for Higher Education Development system. Criteria for such rankings to be considered consistent is then presented, these criteria are applied to rankings in general and to the two above mentioned in particular. Thereupon it is concluded that both rankings show severe methodological limitations and thus are not able to justify their pretense of being reliable methods for evaluating universities.

Key words:

- Higher Education
- Universities
- Evaluation

Introduction

Evaluation can be a powerful stimulus for institutional improvement, but if not done correctly it can also lead to wrong decisions as well as unfair and negative consequences. Evaluation is not an end in itself. It makes sense if it provides information that due to its quality and relevance, may contribute to improvement. Whether its feedback may be used to redirect efforts (formative evaluation) or if it is useful for allocating incentives or sanctions, with due caution (summative evaluation). These considerations are particularly important when the results of performance evaluation and institutional quality are associated with tough decisions, i.e. funding, as has been occurring with the assessments of the so-called university rankings.

Rankings of institutions or programs in general

Publications listing higher education institutions (HEI), or programs, allegedly by quality, are part of the current landscape of higher education systems in many countries. As in many other cases, this phenomenon presented itself earlier in the United States, but at least since the 1990's it has spread worldwide.

Speaking of rankings we will use two definitions. The first one was proposed by the leading scholar of the history of this type of evaluations, David S. Webster, who referred especially to classifications based on institutional prestige, and stressed that each institution had a different place. According to Webster, to make an academic quality ranking, a list “must be ordered by some criterion or set of criteria that the list authors consider to measure or reflect academic quality, and should consist of a listing of the best universities, colleges or departments within a certain area of study, in numerical order according to their supposed quality, so that each school or department has its own place (rank) by itself, and not simply be part of a group among other schools in a few categories, groups or levels” (1986: 5).

The second definition, by Morrison, Magennis and Carey (1995), cited by Bowden, refers in particular to the league tables of institutions that began to be published in the late 1980's in the UK. According to these authors, the tables are based on “weighted combinations of scores generated by performance indicators, in which the overall score is used to sort (rank) institutions such as schools, universities and hospitals” (Bowden, 2000: 42).

Currently, rankings tend to use a combination of “objective” indicators, often of inputs (staff, library, budget) and “subjective” opinions on institutional prestige, so the second definition is better in this respect, but we must not lose sight of the idea in the first definition that a fundamental element of any ranking is that each institution or program occupies a different rank, and is not simply part of a group.

Rankings are not a new phenomenon, although their proliferation is. The idea goes back more than a century in the United States. With a history that dates back to 1888, the idea of classifying institutions based on some quality assessment, was proposed by psychologist James McKeen Cattell in 1910, and was supported by the institutional affiliation of leading scientists (Webster, 1986). According to him, in 1925 Hughes published a ranking based on the opinions of a group of experts, this methodology was later used, with some adjustments, in another ranking system released in 1934 (Webster, 1983). This trend was strengthened after 1959 and, above all, in the 1980's, extending to the undergraduate level. 1983 saw the first edition of *America's Best Colleges* guide, which has been published annually since then by the u.s. *News and World Report* magazine. *Time* and *Newsweek* began publishing similar lists in 1996. According to Rachel Bowden, "an agency that advises institutions on rankings has listed 52 different American publications that rank universities." (2000).

u.s. *News* lists distinguish institutions according to their geographical location and type, based on the *Carnegie Foundation* classification, in addition to Research Universities type I and type II, and significant proportions of undergraduate and graduate students, the lists also distinguish national, regional, and liberal arts institutions.

The place of each institution depends on its situation in relation to indicators of financial and human resources (teachers), the selectivity of its students, its retention and graduation rates, the donations it receives from its alumni and its academic prestige. The weight of each indicator varies depending on the type of institution, but prestige, that is based on personal opinions, usually represents 25% of total weight.

In the UK, the oldest ranking was released in 1992 by *The Times*, and annually thereafter. Since 1998 similar products started appearing in several other newspapers (*Financial Times*, *Sunday Times*) and in special books such as *The PUSH Guide to Which University*, *NatWest* and *Virgin Alternative Guide* (Bowden, 2000).

Bowden also presents the case of an alternative ranking that was published on the Internet only in 1998: the *Red Mole*. This ranking was not based on indicators of resources or prestige among academics or employers, but on the opinion of the students themselves, collected through a voluntary online survey which asked students about the quality of nightlife, boardrooms, teachers and sports facilities.

Although the number of respondents was small (about 1% of total), and did so voluntarily, which is self-selected, the results can not but draw attention, for institutions ranking atop are not the same as those on all other classifications, which in this case fell well below: *Cambridge* 42nd, *Oxford* 35th and *Imperial College* 31st. (Bowden, 2000).

Rankings published in Canada, by *Maclean's* magazine, began to appear in 1990. This classification distinguishes primarily institutions offering doctoral studies, including medical schools, which offer both undergraduate and

graduate (*comprehensive*), and those focused primarily on the undergraduate level (*undergraduate*).

Separate rankings are presented for each of these three categories, based on indicators that are related to students' academic performance, the teacher-student ratio in general or only with *tenure*, teacher quality; the total budget, and the part of it that is devoted to student services and scholarships, the quality of the library, and prestige among alumni, academics and businessmen.

According to a recent count, the number of countries in which higher education institutions are ranked is very large, whether by some media, promoted by the ministries of education or developed by some universities. In addition to the Anglo-Saxon countries already mentioned, to which Australia should be added, the list includes in Europe: Germany, Slovakia, Spain, France, Italy, the Netherlands, Poland, Portugal, Romania, Russia, Sweden, Switzerland and Ukraine; in Asia: China and Hong Kong, South Korea, Japan, Kazakhstan, India, Malaysia, Pakistan and Thailand; and in Africa, Nigeria and Tunisia (Marginson, 2010: 546).

This phenomenon is also present in Latin America, where major newspapers in Argentina, Brazil, Chile, Colombia and other countries issue rankings. In Mexico we have *Reader's Digest* and the newspapers *Reforma* and *El Universal*. The General Directorate of Institutional Assessment at the National Autonomous University of Mexico, on top of analyzing some international rankings (Ordorika and Rodríguez, 2008), develops its own system for classifying Mexican institutions according to their research output, the *Comparative Study of Mexican Universities* (ECUM, acronym in Spanish) (Ordorika *et al.*, 2009, Márquez, 2010).

International rankings

International university rankings also have a relatively ancient history, although those that nowadays are attracting much attention are less than ten years old. Between 1967 and 1983 Jack Gourman published lists in which he sought to classify hundreds of programs, according to quality, some 1,500 u.s. colleges and 700 universities around the world. These rankings, however, did not explain the methodology used and had odd features, such as to distinguish minor differences with hundredths of a point, describing very diverse institutions.

In one case he assigned a score of 4.73 to the Sorbonne, 4.72 to Oxford and 4.71 to the Lomonosov University of Moscow. It is rather unlikely that generally a similar pattern would arise, whereby dozens of institutions would be ordered strictly by a hundredth of a point from one another, with no ties or gaps (Webster, 1985, quoted in Martínez Rizo, 1992: 43).

In the first decade of the twenty-first century international rankings of universities gained unprecedented visibility, probably encouraged by the ease of dissemination over the Internet, particularly in the case of the three presented below, Shanghai, *Times Higher Education Supplement*, and the *Ranking Web of World Universities*.

Shanghai ranking

Since 2003, the Institute of Higher Education at Jiao Tong University, Shanghai, annually publishes the so-called *Academic Ranking of World Universities*. This work receives great attention by the world's media, with increasing impact in terms of decision making by university administrators and ministries of education, as well as policy makers and the public. Despite the harsh criticisms that have been made, its relevance seems only to increase, influencing the educational policies of some countries to the extent that having a certain number of the country's HEIs among the highest rankings becomes an important goal, at system level.

Table 1
Shanghai ranking domains and criteria

Domain	Criteria and weight	Comments
Quality of education	Graduates with Nobel Prize or Fields Medal (10%)	Undergraduate or graduate, if someone studied at several HEIs they all receive points. Weight increased.
Teacher Quality	Scholars with Nobel Prize or Fields Medal (20%)	If someone works at several HEIs they all receive points. Weight increased. Not clear what is meant by "working"
	Highly cited academics (20%)	List of 250 most cited on each of the Thomson Scientific 21 areas
Research productivity	Nature and Science papers (20%)	Within the last 5 years; this doesn't count for social science HEIs, weight varies according to author's range
	Total published papers (20%)	In the Thomson Scientific database, double the weight given to social science
Productivity	Total of 5 previous criteria among fte teachers (10%)	FTE = Full Time Equivalent. This criterion is ignored if data is unavailable

Source: Billaut *et al.*: 2009

For a more detailed description of the *Academic Ranking of World Universities* read Billaut *et al.*, 2009 and Ordorika, Rodriguez *et al.*, 2008.

World University Rankings from The Times Higher Education Supplement (THE)

In an article that presents the most recent changes made to this ranking's methodology, Phil Baty rhetorically asks a question and then provides two different answers:

How do you go about measuring something as intangible as the quality of a

university? The short answer, of course, is that you can't. What can be done, however, what we try to do with these rankings, is to try to capture the most tangible and measurable elements that make a world class modern university (Baty, 2009).

Then he adds:

In 2004, when the THE conceived the first *World University Ranking* along with the signature Quacquarelli Symonds (QS), we identified four “pillars” that form the foundation of any leading international institution, and are beyond discussion: excellent research, high quality teaching, outstanding graduate access to employment and international focus. Much more controversial are the measures used in our rankings and the balance between quantitative and qualitative measurements (Baty, 2009).

The dimensions of the THE ranking concept of quality, how they are measured and their weight in the overall score used for the ranking are summarized as follows:

Table 2
Factors in the Times Higher Education Supplement ranking 2004-2009

Factor	Description	Weight	Comments
Research excellence	Citations from articles published by the institution's staff divided by total faculty	20%	According to Scopus-Elsevier Database
Teaching excellence	Student-teacher <i>ratio</i>	20%	Poor but available
International Focus	% Of teachers from other countries	5%	According to data from each institution
	% Of students from other countries	5%	
Graduate access to employment	Views of academics according to survey	40%	Qualitative worth 50% of tota
	Employer views of graduates according to survey	10%	

Source: Baty, 2009

Above are shown the four dimensions and six indicators used in the THE ranking from 2004 to 2009. The lists include institutions with undergraduate and graduate students as well as covering at least two of five areas: natural sciences, life and health, engineering and information technology, social sciences, arts and humanities. Because of the limitations of the indicators listed above in November 2009 THE announced major changes in the methodology for the construction of its ranking, in addition to no longer working with the company QS (which will continue to publish their listings as QS *World University Rankings*), but with Thomson Reuters, THE will use the ISI Thomson databases for information on indicators of research productivity. THE announced that they have consulted a group of experts in higher education as well as about 40 university rectors, regarding their new methodology.

Factors considered in the preparatory phase appear on the table below, THE announced that the final factors will be made public before the issue of the rankings based on the new methodology, in the fall of 2010 (Baty, 2010).

Table 3
Factors in the ranking of the Times Higher Education Supplement from 2010

Factor	Description	Weight
Economic activity / innovation	Research Income from industry / academia	10%
International diversity	Ratio international students / local	10%
	Ratio international faculty / local	
Institutional Indicators	New entry undergraduate / academics	25%
	Graduate Undergraduate / PhD	
	PhD Graduates	
	Prestige of teaching according to Survey	
	Institutional budget / students	
Research Indicators	Published papers / academics	55%
	Impact according to cites standardized by area	
	Research budget / researchers	
	Research budget external source from government-industry	
	Prestige of research according to survey	

Source: Baty, 2010

The Ranking Web of World Universities

These lists are prepared and published by the Cybermetrics Lab of Spain's National Research Council (CSIC, Consejo Superior de Investigaciones Científicas de España). They are based on the idea that the importance of the Internet is such that it is possible to appreciate how good a university is by analyzing its presence on the web, the production of its faculty and the frequency their respective products are consulted, and in general, its corporate site. Consequently, in this case the unit of analysis is *the institutional web domain*, so that only those universities and research centers with an *independent web domain* are considered. Four indicators were designed from the quantitative results of major search engines (Google, Yahoo, Live Search and Exalead).

Table 4
Factors in the Ranking Web of World Universities

Factor	Description	Weight
Visibility	Total of unique external links received by a site (<i>inlinks</i>).	50%
Size	Number of pages retrieved.	20%
Rich Files	<i>Acrobat and PostScript (.pdf .ps), Word and Powerpoint (.doc and .ppt).</i>	15%
Scholar	Academic papers and citations in <i>Google Scholar</i> .	15%

Source: Website of the Cybermetrics Lab, csic.

The Web Impact Factor (wif) combines the number of inbound external links with the number of web pages in a domain, following a 1:1 ratio between visibility and size. This ratio is used to make the ranking, adding the other two indicators: the number of rich files that a web domain contains, and the number of publications in the *Google Scholar* database. For its relevance to academic activities and publication, as well as volume of use, rich files listed in Table 4 are considered those with .pdf, .ps, .doc and .ppt extensions. This information was taken from this ranking's own website (Aguillo *et al.* 2006 and 2008).

CHE Rankings

The Center for Higher Education Development defines itself as “a reform think tank for higher education”. It was established in May 1994, with funding from the Bertelsmann Foundation and support from the German Rectors' Conference, and is headquartered in Gütersloh, the German Federal State of North Rhine-Westphalia.

CHE's main product of design is a very different ranking from the above, the *CHE Hochschul Ranking* or *CHE University Ranking*. The main features that make it different are summarized below, with information from the CHE page. (<http://ranking.zeit.de>)

The central point is that this is not a simple global ranking but one of detailed analysis, as it explicitly avoids adding partial data to produce an overall score. This decision is based on the conviction that there is no higher education institution that can be considered simply *the best*, since some can be stronger in some ways or in some fields than others. *Instead of crowning an alleged universal winner, we offer a multidimensional system.*

Related to this is the decision not to allocate each institution a single rank, but only to place each one within a group: high, medium or low. This avoids the problem that other rankings have, that the rank of each institution changes substantially from year to year, *by random fluctuations that are interpreted as if they represented real differences.* The differences among institu-

tions in one of three groups and the other two are large enough to be significant, while they are not so within each group, therefore each of these groups is presented in alphabetical order.

Also due to an above mentioned feature, CHE rankings are specific only by area, and not globally, since the quality of institutions in particular disciplines varies greatly, and an aggregate ranking does not provide useful information to guide prospective students nor to take into account institutional decisions to cultivate certain areas with special interest.

Another feature is the methodological diversity, since it uses information from different sources and perspectives, including data on departments and programs, but also views from students, teachers and alumni, which allows us to contrast the views of different stakeholders and is also based on objective data indicators.

The system considers 34 areas of study at the university level ranging from Administration to Social Work, going through Architecture, Biology, Sports Science, Computer, Communication, Law, Economics, Education, Nursing, Physics, Geology, History, Electrical Engineering, Industrial and Mechanics, Mathematics, Mechatronics, Medicine, Psychology, Chemistry and Sociology.

Around 30 numerical indicators are built with data on various aspects raised. These indicators can be used to rank the programs, plus additional information that is not used for the rankings is available to the user of the database. There are 9 general criteria, each includes particular aspects which vary depending on the specific area. The following are some examples of indicators with an approximate total number.

- Labor market and career orientation (20): relation between theory and practice, preparation for work, internships...
- Facilities (18): physical facilities, laboratories, computers, library...
- Research (13): publications, citations...
- Opinions about quality and prestige of the program (4): students, faculty and alumni...
- International orientation (13) foreign students, language requirements...
- Results (10): graduates a year, time to graduate, average grades...
- City and University (26): diversity of choices at the institution, services offered by the host city/town...
- Students (6): entry requirements, proportion of admitted, of women...
- Academic support and teaching (37): class size, student/teacher ratios, teaching methods, individualized counseling, assessment of teaching...

CHE initially considered only ranking institutions and programs in Germany, then it spread to neighboring Germanic-speaking countries, notably Austria and Switzerland. Since 1998 it included more than 130 universities and over 150 universities of applied sciences (Fachhochschulen). The latest version considers the opinions of some 200,000 students and 15,000 teachers (<http://ranking.zeit.de>).

CHE results are published by the newspaper *Die Zeit*, but not in the form of lists as in the case of other rankings, but consistent with what was stated before, using an interactive online system that allows each user (future students or others) to customize the criteria for sorting the options based on their own interests (Marginson, 2010: 547).

From this experience CHE, along with the Center for Higher Education Policy Studies at Twente University (Netherlands) and others, are developing a ranking with a similar concept and greater international reach, the *U-Multirank*, currently being piloted in 150 institutions around the United States and Canada, Australia, Asia, Africa and Latin America.

Towards an assessment of the rankings

In some cases rankings are produced by specialized institutions, in relation to a communication medium, as is the ranking of the CHE and *Die Zeit*. Sometimes they are made by academic institutions such as *Shanghai* ranking or the *Ranking Web of World Universities* by the Cybermetrics Lab of the Council for Scientific Research in Spain.

In many cases, however, rankings are produced without any qualified support by some media supposedly motivated by a claim of contributing to increase the information available regarding the higher education market, providing elements to parents and prospective students to choose an institution or program. There is also a clear intention of increasing the respective newspaper's or magazine's sales, given the growing interest in rankings. In fact, the media that publish rankings get juicy sales each year, as the season when young people finishing high school must decide which programs and institutions to apply for, which for many represents the first important decision with life implications.

But above any of these more or less legitimate interests, the fundamental questions to answer so as to make an opinion about rankings should be related to their quality, since judgments about it will be very different whether or not the information they provide is good.

As with any evaluation, a ranking involves identifying what is being evaluated; defining what constitutes quality and clarifying the reference to contrast the condition of those being evaluated; to make the concept of quality operational and defining the dimensions and indicators to measure reliably, contrasting the result of measurement to the reference and making sound value judgments. Specifically, for a ranking to be considered good quality, it should address the following issues:

- Clear identification of *what is being evaluated (valuees)*: institutions, units or programs to be evaluated.
- Precise definition of the concept of *quality* that is expected from valuees. How can they be considered good, specifying their dimensions and indicators for each dimension, making it possible to assess whether the rankings are *valid*.

- Quality of information used for each indicator, identifying the sources and ways to obtain this information in order to assess whether the measurements are *reliable* and whether, based on them, if consistent comparisons about the evaluatees can be made.
- Solid justification for the means by which indicator measurements are added so as to generate scores on a single scale to place evaluatees in order, to determine whether to conceive quality as unidimensional is justified and done consistently.
- Sufficient *accuracy of measurements*, so that it is possible to see if the difference that distinguishes evaluatees from each other is significant and consistent, or not, so that it makes sense to rank distinguishing one by one the places evaluatees occupy in the ranking.
- Identification of *reference standards*, for comparisons to be made to assess whether an evaluatee can be considered, or not, good quality.
- Consistency of emerging trends and results.
- Correct formulation of the judgments reached by the very fact of making a ranking, cautiously taking into account the context of institutions being compared to assure fairness, and how the consequences, intended or not, of the ranking have been taken into account.

The first seven points listed above coincided in their technical nature, because if the rankings are elaborated poorly, so that they make wrong measurements of what they intend to evaluate, or measure something different, regardless of good intentions, the results will not be able to firmly support appropriate uses.

But even if measurements are accurate, the results can be used properly or improperly, so it is important to consider the eighth and last item, about the contextualization of value judgments and applications of the results. The choice of these eight criteria to assess the quality of a ranking is justified in part based on the literature of the field, which in different ways, coincides in identifying as dimensions of all good assessments, the validity and reliability of the measurements, on which the judgment is based, its unidimensionality to justify aggregating data from different indicators, its accuracy and consistency over time (e.g. Bird, 2005, Goldstein and Spiegelhalter 1996; Shavelson, McDonnell and Oakes, 1989). On the other hand, the author is involved in the development of a system of indicators by the National Institute for the Evaluation of Education, and has thus systematized these criteria (Martínez Rizo, 2005, 2007a and 2007b).

In the following paragraphs the eight criteria above will be applied to the overall rankings.

Definition of the object to evaluate

The problem with rankings of complete institutions start with the definition of their subject. Universities are large and complex, and their quality is essentially multidimensional. They have strengths and weaknesses, may excel in research yet teaching may be weak, alongside very productive research

groups may be other mediocre ones, some programs with excellent graduates and others stagnant with poor and outdated training.

Underlying concept of quality, dimensions and indicators

Assuming that the evalee is well-defined, the following step is related to the concept of quality, the indicators that operationalize it and the information that supports them. Usually there is information on some aspects of quality and not on others, and often, the most abundant data concerns less important aspects, while information related to fundamental aspects is scarce or of questionable quality. Which tend to happen in many rankings.

Quality of measurements and of information used

In assessing a ranking's quality two issues ought to be discerned: First, if it is based on facts or opinions; and second, in either case, is its information solid. The oldest rankings were based on the quality of the faculty, which is undoubtedly central to the quality of an institution or program, teaching or research, but given there was no objective information about the achievements of teachers or researchers, institutions were evaluated based on the views that some scholars had of the evaluated institution's academics. Therefore these were called reputational rankings.

The limitations of this type of approaches are obvious. Even in a higher education system with burgeoning mobility, information and communication, as is the case for the United States, there are obvious risks of high subjectivity, lack of references and a "halo effect" in the opinion of teachers, businessmen or others on the quality of institutions or programs. Webster cites the case of a ranking of undergraduate schools in the area of business, which some experts mentioned among the best universities: Harvard, Yale and other highly praised institutions that did not happen to have undergraduate programs in that area at the time. In Mexico, the National Autonomous University of Mexico (UNAM) and Instituto Tecnológico y de Estudios Superiores de Monterrey (ITESM) are often considered as *institutions of excellence* in all studies based on opinions, but it is clear that, beyond the actual quality these and each of their areas have, the reasons for them being repeatedly mentioned has to do with the fact that these institutions are much more visible than others. It is symptomatic that institutions such as the Centro de Estudios Avanzados (CINVESTAV) and El Colegio de México are hardly mentioned in these studies, although they excel in some areas not only nationally, but internationally.

Today, as we have seen, rankings tend to combine subjective views on the prestige and objective data, overall institutional resources or inputs that can be obtained easily, such as the student-teacher *ratio*, the proportion of full time or part time teachers; of staff with doctorates or without them, the number of volumes in the library, or the minimum score that an applicant must obtain in the admission test.

These more objective data appear more robust than those based on subjective opinions, but have the problem that they do not measure the quality of the institution or program, but only measure their wealth and resource endowment. Certainly there is a link between the abundance or scarcity of resources and the quality of an institution or program, but that relationship is not unambiguous nor linear. There are well-endowed institutions, mediocre or downright bad, and there are modest and excellent, and to distinguish one from another is not enough to measure their inputs. In the case of teaching, for example, to see if institutional resources, few or plenty, are used efficiently, reliable data is needed on graduates' competence level, after subtracting the level they had before entering the institution and taking into account dropouts. All these data are not currently available, of course, neither it is likely that they will be on a scale beyond that of a few rather complex research projects.

Justification for the factor aggregation component and its methodology

A basic methodological principle is that the dimensionality of measurements relating to any latent construct, such as quality, must be verified. Aggregating data from multiple indicators can only be justified if they all relate to the same domain. In that case there are several ways of accomplishing this, all with pros and cons, but none is as simplistic and naive as those used in many rankings, with the aggravation that indicators from clearly distinct dimensions are often added, such as quality of research and teaching, or different and distinct units within the same institution.

Measurement Accuracy

Ranking reviews often show that the authors do not usually have a good grounding in measurement, nor seek advice thereon. It is no wonder that the procedures for quantifying each of the indicators, which are used to add the partial results, and build the single rank scale, are done without respecting basic methodological rules. This failure is reflected, *a fortiori*, in the doubtful accuracy of their measurements, as shown when no justification is offered based on which to defend why an institution would have a lower score than another, despite their differences, that would involve a corresponding differentiation in quality. No one knows the margin of error that inevitably results, and it is therefore impossible to assess the significance of these discrepancies.

Definition of reference standards

Obviously this type of definition is not available, but it is indispensable for the value judgments that are reached to be not only relative but absolute.

Rankings can only reach relative judgments, in the sense that they claim an evaluatee to be *better or worse* than others, without specifying, as noted above, if it is significantly better or worse, or only slightly, while the differences are not significant.

Yet being better than something/someone is not necessarily positive, or being worse than something/someone is not necessarily negative. No university would consider it bad to be a little worse, or less good than Harvard, nor look upon being a little better than a clearly deficient institution as positive. Everything depends on the point of comparison, which should not be simply the situation of other *evaluatees*, but a more robust parameter, which rankings typically lack (ORPEE, 2010).

Consistency of trends and results

As a result of the deficiencies referred to above, rankings of universities and programs often present major changes in the places that some of the evaluatees occupy, with no basis to believe that these changes reflect actual changes in quality. Such institutions do not easily change from year to year, unless dramatic reforms or disasters occur. Thus the existence of abrupt changes should be considered a clear indication of the rankings lack of consistency.

Contextualization and consequences: the uses of the results

Due to their strong media impact, tough decisions are increasingly associated with the rankings results. Moreover, the underlying measurements often lack the qualities of validity, reliability and comparability which are essential to support objective value judgments about the quality of the institutions being evaluated. The conclusion is that the use of such rankings will easily lead to committing grave injustices by failing to truly identify the institutions or programs that should really stand out for their quality and deserve encouragement; those that, despite not yet having excellent results are making commendable efforts from a disadvantaged situation and may need support to take off; and those exhibiting severe repeated and unjustified failures that would justify remedial action.

An assessment of the Shanghai and the Times Higher Education Supplement rankings

These rankings exemplify poor methodology, combined with strong media impact, leading to applications that cannot aid in decisions that lead to real quality improvements. A brief review of the above mentioned methodological criteria will help to appreciate that the two most visible rankings fail to meet them.

Failing to define the objects they are to evaluate: As was previously mentioned a university is a complex whole, and therefore it is not reasonable

to reduce its merit to a single place on a one-dimensional list. Moreover, in the case of the Shanghai and the THE rankings it is not clear what should be considered a university nor how the list of those considered in the ranking is prepared. How to define if Berkeley and UCLA, with eight other institutions that form the *system* of the University of California, should be considered separately or as a whole? The same question arises in the case of UNAM's branch campuses, the different units of the Universidad Autónoma Metropolitana and Univesidad de Guadalajara or the different Tecnológico de Monterrey campuses, or the universities resulting from the division of the University of Paris. In terms of Mexican institutions it is interesting to note that in the opinion polls among scholars and employers which constituted 50% of the total weight of the THE ranking those consulted were only asked about UNAM, ITESM and the Universidad Iberoamericana. No other institution was represented in the consultation.

Assuming a quality concept that is partial and operationalized with a bias: Apparently the term "world class university" refers to one that produces more research, but since this is only one of three or four basic functions of higher education institutions it is legitimate to ask why is it so privileged by these two rankings. Moreover, the dimensions and indicators used are obviously very strongly biased in favor of certain fields of knowledge and certain geographic and linguistic regions.

Poorly measured and inadequate handling of information: Although all the indicators used are of the objective type, the measurements are very poor, due to the limitations of the database from which information is taken, as well as unjustified decisions on how to attribute the production of a Nobel prize and other leading scientists to a particular institution, establishing the elapsed time from when the papers taken into account were published, among other reasons.

Arbitrary and deficient adding of factors: A notorious weak point when considering that the top five factors overall, not relative to size, make the largest institutions have a significant advantage by this fact alone. This is not later corrected by using this criterion as the sixth factor, but moreover that this last factor departs from the aggregation of the previous ones, without justifying the weight assigned to each of them, is methodologically unacceptable.

Very poor measurement accuracy: No justification is offered based on why one institution gets a lower score than another, regardless of their differences, necessarily implying a corresponding difference in quality. There is no respect for the basic methodological principles of any measurement, the results' margin of error is unknown and therefore it is impossible to assess the significance of these differences.

Lack of definition of reference standards: Neither of these rankings in question explains if some values of the scale used can be interpreted as good or bad, which would allow to go beyond knowing whether an institution is better or worse than another or others, as this is not necessarily in itself positive or negative.

Consistency of trends and results: In addition, according to these rankings, some institutions could be much better than others one year and much worse the next. Some examples will suffice to demonstrate the weaknesses in this regard.

In the Shanghai ranking, the University of Buenos Aires in 2004 was ranked 295 and 279 in 2005, an improvement of 16 places in one year is pretty amazing, but in 2006 it peaked at 159, a jump of 120 places is simply not credible. Nor is the decline of 58 seats suffered by the Autonomous University of Madrid between 2005 and 2006, going from 198 to 256 (Ordorika, Rodríguez *et al.*, 2009: 28). As for the THE ranking, UNAM was placed 195 in 2004, rose to 95 in 2005 and 74 in 2006 but then dropped to 192 in 2007, climbing to 150 in 2008 and down again to 190 in 2009 (Ordorika and Rodríguez, 2010: 20).

Contextualization and consequences: How the results are used. International rankings are drawing great attention from the media, to the extent that major decisions tend to rely on them, the inequity that their lack of reliability implies, both for institutions unfairly penalized as well as unfounded beneficiaries. The severity of the negative impact of these rankings can be seen when considering that institutions and ministries of education in some countries tend to base their policies on their results. Some examples can be cited: the University of Guanajuato, Mexico, in its recently approved Development Plan aims to reach to be *among the top 100*, and in France the government's announcement that, in the near future, the country aims to have four institutions in the top 20.

Other analyses have reached similar conclusions. Billaut notes that the first serious analysis of the Shanghai ranking, done in 2005 (Van Raan, 2005a and 2005b and Liu, Cheng and Liu, 2005) concluded that it should not be treated as a true assessment, and that the most serious problem with the rankings is that they are an example of "what is absolutely unacceptable" (Billaut *et al.*, 2009: 33). Using the Multiple Criteria Decision Making approach, a group led by the author revised the criteria and the aggregation method used for the Shanghai ranking. His main conclusion is that the criteria are irrelevant and that the aggregation method has serious problems, so the ranking "is not an appropriate tool to analyze the quality of institutions" (Billaut *et al.*, 2009). From these authors, two paragraphs that illustrate eloquently the consequences that can take an uncritical use of these listings:

Apparently, the Shanghai *ranking* aims to answer the question of what is the best university in the world. For some readers the question itself may seem childish and without much interest. We agree. However, those readers should be aware that there may be lazy decision-makers who can simply use the results of a ranking that is within their reach. And most importantly, strategic decision makers who can use the results to promote their own ideas about how to reorganize a higher education system. Moreover, as with any management tool, the existence of a ranking helps to change the behavior of the agents involved, sometimes producing undesirable changes (2009: 2).

Suppose you manage a university and want to improve its position in the ranking. It's pretty simple. There are important areas of your school that do not add to the position in the ranking, such as law, humanities and most social sciences. Eliminate those areas, probably save a lot of money. Use it to form research groups that do help improve your position. Some of Thomson Scientific indexes are quite useful for that purpose, after all, the list of the probable next five Nobel prizes in medicine is not as long and, anyway, if your hires do not receive the Nobel, they will certainly publish a lot in refereed journals that count in the ranking, and most likely will be on the list of highly cited researchers. This tends to promote a way of seeing science that is a lot like professional sports where a few rich teams compete to attract the best players in the world. We're not quite convinced that this is the best way to increase human knowledge, to say the least (2009: 32).

As for the *Times Higher Education Supplement* ranking, more than ten years ago an analysis listed the following points, as only some of the methodological challenges that have been posed against rankings and league tables:

- Technical status of some variables;
- Inadequate construct validity
- Scaling the variables;
- Changes from one year to another in the variables and their weights;
- Data manipulation;
- Inconsistency of changes;
- Mismatch between the global score and the quality of each academic unit, and
- Distortion of the institutional purpose (Yorke, 1997: 62).

Conclusion

The analysis of the criteria used by some of the most important rankings of universities internationally, and the coincidences with similar analysis, leads to the conclusion that the serious methodological flaws they have make them highly inappropriate tools for reliably assessing the overall quality of the institutions they claim to evaluate.

Even limiting their influence to the less ambitious pretense of evaluating only the quality of the research function, their deficiencies are so severe that not even with those limits can the results be reliable and are still not a good basis to ground decisions by authorities, which can lead to errors that may prove too costly for institutions.

Contrary to what the media that sponsor many of these initiatives pretend, nor will the rankings of individual programs be an appropriate guide for future students taking on the task of choosing the institution and program that they will seek to be admitted into. Taking into account that there is no better university or best program *at all*, but there are programs more or less suitable for certain applicants, we will understand that the most popular rankings are no substitute for a good career guidance system. The rankings

produced by the CHE are much better for that role, due to the fact that they have less ambitious claims and suitable features for such purposes.

No methodology can properly assess, in one single dimension, the quality of institutions which, in a global sense, is essentially multidimensional. It is more feasible to evaluate units, quality functions or aspects of a university that can reasonably be defined as one-dimensional. Assuming an appropriate methodology, it does not seem absurd to compare different, for example, programs in medicine or social work, doctorates in astronomy or anthropology, or research groups on lung cancer or on methodologies for teaching reading.

In the absence of evaluations there is a tendency to treat all institutions alike, it does not seem right. But based on inadequate evaluations something worse happens: the differences in treatment have nothing to do with objective, but illusory merits. The result may be that good things would tend to deteriorate, and things needing consolidation would not mature and deficiencies would prevail and even get worse.

Good quality evaluations of higher education institutions are not impossible, but involve approaches that would not fall into the reductionistic simplifications of the common rankings. Only with an array of complementary formulations could a reasonably comprehensive approach be attained to something as complex as the quality of a university.

References

- Aguillo, I. F., J. L. Ortega y M. Fernández (2008). “Webometric Ranking of World Universities: Introduction, Methodology, and Future Developments”, en *Higher Education in Europe*, Vol. 33 N° 2-3: 234-244.
- Aguillo, I. F., B. Granadino, J. L. Ortega y J. A. Prieto (2006). “Scientific research activity and communication measured with cybermetric indicators” en *Journal of the American Society of Information Science and Technology*, Vol. 57, N° 10: 1296-1302.
- Baty, Phil (2010). *The unveils broad, rigorous new rankings methodology*. June 3. Consulta en <http://www.timeshighereducation.co.uk/storyasp?sectioncode=26&storycode=411907&c=1> en 2010-08-10.
- Baty, Phil (2009). *Rankings 09: Talking Points. The 2009 world ranking methodology*. October 8. Consulta en <http://www.timeshighereducation.co.uk/storyasp?storycode=408562> en 2010-08-10.
- Billaut, Jean-Charles, D. Bouyssou y Ph. Vincke (2009). Should you believe in the Shanghai ranking? *A Multiple Criteria Decision Making view*. 29 May.
- Bird, Sheila M. et al. (2005). “Performance indicators: good, bad and ugly”, en *Journal of the Royal Statistical Society A*, Vol. 168, Part 1, pp. 1-27.
- Bowden, Rachel (2000). “Fantasy Higher Education: university and college league tables”, en *Quality in Higher Education*, Vol. 6, N° 1, pp. 41-60.
- Center For Higher Education Development (CHE) *Hochschul Ranking o University Ranking* <http://ranking.zeit.de> Consulta en 2010-08-10
- Goldstein, Harvey y D. J. Spiegelhalter (1996). “League tables ad their limitations: statistical issues in comparisons of institutional performance”, en *Journal of the Royal Statistical Society A*, Vol. 159: 385-443.
- Liu, N. C., Y. Cheng y L. Liu (2005). “Academic ranking of world universities using scientometrics: A comment to the ‘fatal attraction’”, en *Scientometrics*, 64: 101-109.
- Marginson, Simon (2010). “National and International Rankings of Higher Education”, en Peterson, P., E. Baker y B. McGaw, *International Encyclopedia of Education*, 3rd Ed. Amsterdam, Elsevier-Academic

- Press. Vol. 4, pp. 546-553.
- Márquez J., Alejandro (2010). "Estudio comparativo de universidades mexicanas (ECUM): otra mirada a la realidad universitaria", en *Revista Iberoamericana de Educación Superior (RIES)*. México, IISUE-UNAM/Universia, Vol. 1 N° 1 pp. 148-156. <http://ries.universia.net>
- Martínez Rizo, Felipe et al. (2007a). *Propuestas y experiencias para desarrollar un sistema de indicadores educativos*. México. INEE.
- Martínez Rizo, Felipe (2007b). "Propuesta metodológica para desarrollar un sistema de indicadores para evaluar la calidad de la educación en México", en Varios. *Seminario Internacional de Indicadores Educativos*. Memoria. México. INEE. Julio 9.
- Martínez Rizo, Felipe (2005). "El diseño de sistemas de indicadores educativos: consideraciones teórico-metodológicas", en *Cuadernos de Investigación*, N° 14. México, INEE.
- Martínez Rizo, Felipe (1992). *La calidad de las instituciones de educación superior. Su evaluación y su promoción*. Cuadernos de Planeación Universitaria, 3ª época. Año 6, N° 2. México. UNAM.
- Morrison, H. G., S. P. Magennis y L. J. Carey (1995). "Performance indicators and league tables: a call for standards", en *Higher Education Quarterly*, 49(2), pp. 128-45.
- Observatorio Regional de Políticas de Evaluación Educativa (2010). *Los rankings de escuelas como forma de entregar información sobre calidad educativa*. Boletín monográfico N° 17, agosto, 11 págs.
- Ordorika, Imanol y R. Rodríguez Gómez (2010). "El ranking Times en el mercado del prestigio universitario", en *Perfiles Educativos*, Vol. xxxii, N° 129, pp. 8-25.
- Ordorika, Imanol, et al. (2009). Desempeño de universidades mexicanas en la función de investigación: estudio comparativo. Datos básicos 2007. *Cuadernos de Trabajo de la Dirección General de Evaluación Institucional*, N° 1. UNAM-DGEI, 37 págs.
- Ordorika, Imanol, R. Rodríguez Gómez et al. (2008). Comentarios al *Academic Ranking of World Universities 2008*. *Cuadernos de Trabajo de la Dirección General de Evaluación Institucional*, N° 2. UNAM-DGEI, 97 págs.
- Shavelson, Richard J., L. McDonell y J. Oakes, Eds. (1989). *Indicators for Monitoring Mathematics and Science Education. A Sourcebook*. Santa Monica, Rand.
- Van Raan, A. F. J. (2005a). "Fatal attraction: Ranking of universities by bibliometric methods", en *Scientometrics*, 62: 133-145.
- Van Raan, A. F. J. (2005b). "Reply to the comments of Liu et al.", en *Scientometrics*, 64: 111-112.
- Webster, David S. (1986). *Academic Quality Rankings of American Colleges and Universities*. Springfield, III, Charles C. Thomas Publ.
- Webster, David S. (1985). "How not to rank universities", en *Higher Education*, Vol. 14, N° 1, pp. 71-79.
- Webster, David S. (1983). "America's Higher-Ranked Graduate Schools, 1925-1982", en *Change*, May-June, pp. 13-24.
- Yorke, M. (1997). "A good league guide?", en *Quality Assurance in Education*, 5(2), pp. 61-72.