

Summary Report of the Invitational Roundtable on Statistical Indicators for the Quality Assessment of Higher/Tertiary Education Institutions: Ranking and League Table Methodologies

JAMIE P. MERISOTIS

The number of ranking and league table systems has increased rapidly in recent years across many nations. However, little cross-national analysis of the strengths and weaknesses of various approaches has been conducted. To accomplish such a task, a framework or typology that describes the individual ranking systems is needed. This article presents the summary principles from the first-ever international meeting on these issues, held from 13 to 15 June 2002, in Warsaw, Poland. These principles will be of assistance in future discussions about the different benefits and limitations of ranking and league table systems.

THE CONTEXT OF RANKING SYSTEMS

It became apparent from the papers presented and the discussions undertaken during this historic meeting that a foundation must be established to define the context of the issues of ranking and league table systems. The number of systems developed in recent years has increased rapidly. The conference presented about ten or so different approaches, which is a small sampling of the numbers of approaches that have been developed in many countries. The *US News and World Report* annual college rankings issue appears to have been a precursor that has spawned many different approaches.

This proliferation of ranking systems is not surprising. With the massification of higher education around the globe, there is an increasing appetite for independent information on higher education. The private sector has responded to this increased appetite with apparent commercial success, since without a consumer market the rankings would not be published.

FUNDAMENTALS OF RANKING SYSTEMS

Who Ranks?

Certain fundamentals of ranking systems were discussed throughout the meeting. One of the basic fundamentals of ranking is *Who ranks?* The assumption is that the ranking taking place is media- or private publication-oriented, but there are additional rankings being conducted in at least two other sectors. The first is that of professional associations that operate ranking systems and develop unique approaches. Governments constitute the other sector that plays a role in rankings. The private, media-based rankings are the dominant, but not exclusive, developers of ranking systems.

Why Rank?

The second fundamental question is *Why rank?* A primary purpose is to provide consumer information to help people make choices about higher education. Institutional marketing

purposes and assistance to the institutions themselves so as to help them distinguish what they are and what they are doing for consumers of higher education constitute another motivation. A third motive for ranking is, indirectly, to have an impact on quality, primarily because the rankings have the effect of promoting competition and therefore influence.

The Audience for Rankings

The *audience for rankings* is the third fundamental. Obviously, students are a major consumer. Parents are an important, and in some ways superior, group targeted by ranking systems, particularly in cases in which tuition fees are involved. Parents play a large role in the decision-making process. Additional audiences for rankings include the institutions themselves, as well as employers and government.

The Ranking Process

The final fundamental is the *ranking process*. While this process varies considerably according to the system or approach being utilized, it generally includes a logical set of elements. The first element is the collection of data. The data are either mined from existing data sources or are collected from original sources specifically for the ranking or league table. The second step is selecting the type and quantity of variables from the information collected. Once that task has been completed, the standardizing of indicators from the variables takes place, and the weighting of the various indicators is determined, followed by the physical process of conducting the calculations. Somewhere between that point and the actual recording of the results, “adjustments” are made, in some cases, to the rankings for a variety of reasons. Some of these reasons may be political, which can be: “This list doesn’t look right. It doesn’t look like it has the right institutions.” Another example concerns more technical issues of problems with certain elements of data, with perhaps some of those elements needing to be discarded in order to yield a fair or appropriate ranking system.

A FRAMEWORK FOR RANKING SYSTEMS

A framework or a typology of ranking systems is useful in the design of future conversations and meetings. This first step is necessary for the future organization of a series of effective international conversations about these issues.

Types of Rankings

A basic component of ranking systems is the *type of ranking* that is being used. The many different approaches to ranking can be divided into three categories.

The first category is the unified, aggregated, or single score ranking of institutions. Such rankings rank entire institutions and attempt to pull together a disparate set of indicators, go through the calculation process, apply weights, etc., and draw a conclusion about the overall quality of a given institution. This focus or purpose has been characteristic of most ranking systems. Examples include the *Perspektywy* magazine approach in Poland and the *US News and World Report* rankings.

The second of the three types consists of rankings based on discipline, programme of study, or subject area. These method ranks institutions according to the specific programmes or subjects that are offered. It can cover the varying levels of higher education from

undergraduate or first degree, to graduate, to professional programmes, as well as others. The *Perspectywy* approach is also discipline based, as is the *CHE-Stern* system of Germany, *Business Week*, the *Financial Times*, and many others. The unified and the discipline based are the two main types of rankings.

The third type can simply be called “other” because of the tremendous variations that are possible. Rankings that cannot be easily categorized are included in this category. A good example is the *Recruit Ltd.* approach of Japan, which ranks institutions merely by each of the eighty-eight questions in its survey. The entire ranking system simply reports Question 1, followed by the rank order of institutions, then Question 2 with the rank, etc. There is no other sorting mechanism built into the system.

Many systems have varying combinations of these approaches. One example is a unified ranking system that also undertakes separate rankings by discipline.

Structure

The second consideration is the *structure* of the ranking. One structure is numerical rankings. This approach is the 1, 2, 3, 4, 5 one that is seen in many examples, from *The Times Higher Education Supplement* in the United Kingdom, to the Nigerian example, and to the DAAD approach in Germany. The clustering or grouping of rankings is another structure. The *CHE-Stern* approach with top, middle, and bottom tiers is a good example of this clustering or grouping approach. Another category of structure, which can be referred to as the top-level approach, is ranking institutions numerically but reporting only a fixed number at the top. This is a sort of hybrid or variant of the two other approaches. *Recruit Ltd.* engages in this practice, with only the top fifty surveys listed for each question. A variety of other combinations of these structures is possible as well.

Frequency

The third component in a typology deals with the *frequency* of the ranking. Whether it be annual, biennial, triennial, or some irregular interval, there has to be some focus on the frequency of the ranking in the structure.

Sorting

A *sorting* of the institutional rankings is another necessary component. A common method of sorting is based on institutional control, *i.e.*, whether an institution is publicly or privately owned and operated. The age of the university or the geographic distribution of universities, as in the Nigerian approach, are other methods of sorting. Sorting by mission or focus of institution is also common.

Data Sources

The final component in the typology is the *source of data*. This component is probably the most difficult part to constitute of any kind of typology. Trying to document the sources of ranking systems can become unwieldy when undertaken at an international level, because of their tremendous number and diversity. The strategies adopted by the publishers of these ranking systems generally reflect some kind of combination of mining of existing data and original data collection. Some examples of the sources to include in the typology would be data reported by institutions for certain prior purposes, such as enrollment, funding,

facilities, or other purposes. Additional sources are student surveys, attitudinal surveys, outcomes surveys, employer surveys, analyses of public records concerning unemployment, independent evaluations, faculty and staff surveys, graduate or alumni surveys, accreditation information, etc. All of these information sources would need to be included in the “sources of data” section of the ranking typology.

To restate, the main components that must be included in a typology are the following:

- type;
- structure;
- frequency;
- sorting mechanism;
- sources of data.

SUMMARY PRINCIPLES

- (i) *An independent analysis of what higher education is and does is valuable and important.* The above statement reflects the first principle or conclusion to be drawn from the meeting. The relevant information can be conveyed, through appropriate methodologies, by a ranking system. Ranking systems are likely to be continuing enterprises, and cannot be ignored or discarded. With the appropriate methodologies, they can help to provide effective independent analysis of what higher education is and does. Independent analysis is important, for self-ranking is not feasible and would not be accepted by consumers. Public demand for sorting, given the increase in massification, is simply going to continue to feed the desire for ranking systems.
- (ii) *Ranking systems should be one of a diverse number of approaches to assessing higher education inputs, processes, and outputs.* This second principle or conclusion suggests that rankings can provide comparative information and improved understanding of higher education, but that no ranking system can accurately measure quality. In fact, it would seem that none of the systems presented during the meeting really measures pure academic quality except perhaps for the components dealing with teaching assessments. Most are indirect measures of quality. Therefore, no ranking system should be used as the primary or sole mechanism for assessing the major inputs, processes, or outputs of higher education.
- (iii) *No methodology is infallible or immune to revision or reassessment.* The third conclusion is that various methodologies and approaches have strengths and weaknesses, but that no methodology is infallible or exempt from revision or reassessment. The ultimate decision about which methodology is the right one has to be culturally specific, and grounded in the unique educational, political, economic, and social context of a nation, region, or system. Yet these methodologies themselves are not sacrosanct. There is no magic to them. They are developed by people who are prone to error, as are all human beings. Therefore, revision and reassessment is important and appropriate.
- (iv) *Transparency is essential to the success of any ranking system.* Transparency in how methodologies work and what they are intended to show is critical to the credibility and usefulness of ranking systems. How the ranking is done should be clear and unambiguous. Lack of transparency on the part of the organization that does the ranking can be a mistake because it erodes consumer confidence and fosters mistrust and suspicion among institutions.
- (v) *Integrating the results of a series of rankings into a single score for an institution*

seems problematic and statistically unreliable. The above conclusion is inescapable. Compelling evidence presented at the meeting indicates that the single score idea is problematic for ranking systems. A serious examination is necessary of what information an aggregated or unified ranking system is really conveying from a statistical perspective and a public service or consumer information perspective. What is the story that these aggregated systems are telling? Are they simply conveying results that claim to be objective, but are in fact statistical artifacts? Related to this matter is the principle that it is important to make an effort to avoid annual or at least frequent changes to the methodologies or formulae, because doing so creates problems in terms of comparability across years.

- (vi) *Rankings that cluster institutions into bands or tiers are a more appropriate and methodologically sensible approach than numerical rankings.* There is compelling evidence for this additional statement. These types of approaches help to avoid misinterpretation of small differences in indicators. They reduce the impact of the so-called interval effect. “What really is the difference between numbers 17 and 22 on the ranking scale?” While there is still an interval effect asking “What is the difference between the bottom of the top tier and the top of the middle tier?”, it is reduced using a clustering approach. Consumers like numerical rankings. They like to see who is number 1, yet, there has been success in systems that avoid strict numerical rankings. The *Michelin Guide* approach to clustering for airlines, or restaurants, or cars is a more useful and productive methodology for the development of ranking systems in the future.
- (vii) *Consumers need to have more control and ownership over what is actually being ranked.* A powerful example is the *CHE-Stern* ranking system, which permits consumers to use technology to select which features they think are the most important in ranking a university. The consumer can create a ranking based on his or her own selected criteria. The simplicity of such an approach is also impressive. The required database is not a complicated one to develop. It simply requires some structure to allow it to do calculations. It gives the students or other consumers—such as employers, parents, or the institutions themselves—a chance to do their own ranking of the highest priorities they have in selecting a college or a university, rather than having some external entity dictate to them the most important characteristics in determining the quality of a higher education institution. That approach would go a long way towards improving the validity and utility of these ranking systems.

ISSUES FOR FUTURE CONSIDERATION

It is important to define several broad issues for future consideration, along with some specific future questions.

Continuing Work on the Development of a Typology or Framework

Such work would be very important for international conversations and for learning and sharing across countries. To continue to engage in this cross-national dialogue, it is necessary to establish standards. Otherwise, terminology will be used and a structure or approach talked about, but what is meant in one country will mean something different in another country, and misunderstandings might develop.

Evaluation of Specific Methodologies

Once the framework or typology has been developed, systematic analysis and evaluation of specific methodologies should be continued while keeping the culturally specific context in mind. Currently, the amount of real evaluative work is quite limited. A forensic analysis is needed to understand the specific advantages and disadvantages of the various approaches that are in place. The people who are responsible for these ranking systems can then be urged to make their own judgments as to what is the best approach or methodology.

Are There Core Indicators that are Consistent across Several Countries?

It would be useful to determine whether there are a limited number of indicators that are used in many countries that would be useful for cross-national comparative purposes. From a research perspective, it is important to engage in that process.

E-Learning or Distance Learning?

An additional item for future discussion is that of how ranking systems can or should deal with E-learning or distance learning or the sort of hybrid teaching and learning approaches that are being developed on many university campuses. This development presents a challenge to the traditional approaches of ranking systems.

Distinguishing between or among Undergraduate or First Degree Students and Students at Other Levels

In the case of unified institutional rankings, a key question is how or whether these rankings distinguish between or among undergraduate or first degree students and students at other levels. Many ranking systems blur those distinctions or meld them all together. In some cases, attempts are made to disaggregate the various levels. The strengths and weaknesses of the unified approach related to this issue need to be addressed.

Inter-Disciplinary Studies, Sub-Disciplines, or Programmatic Specializations in the Ranking Systems

Finally, in the case of disciplines or specific subject areas, how should the approaches deal with inter-disciplinary studies, sub-disciplines, or programmatic specializations in ranking systems? The neat categories of discipline, *i.e.*, chemistry, biology, or economics, are no longer effective on university campuses.

CONCLUSION

As the first-ever international meeting dealing with higher education ranking systems, much important information was gleaned from a variety of national and regional contexts. Clearly, more research and continued dialogue are needed to better understand how ranking systems work, what value they contribute to higher education, and how they might be modified or improved to address the evolving nature of higher education. These future conversations will look back upon this historic meeting as the starting point of much-needed cross-national dialogue and resource sharing.