Big is (made) Beautiful
Some comments about the Shanghai ranking of world-class universities

Michel Zitt[1/2], Ghislaine Filliatreau[1]

[1]*Observatoire des Sciences et des Techniques (OST), Paris (France)*
[2]*INRA-Lereco, Nantes (France)*

Revised version of the paper presented at the 1st Intl Conf on World Class Universities, Shanghai, June 2005.
Final version in Jan Sadlak and Nian Cai Liu, Eds. *The World-Class University and Ranking: Aiming Beyond Status,* UNESCO-CEPES/ Cluj University Press, Romania; Institute of Higher Education, Shanghai Jiao Tong University, China, pp. 141-160.

http://www.cepes.ro/publications/WCU/contents.htm

1 Introduction

This study will focus on some methodological questions raised by university assessment, and in particular by the recent "Shanghai ranking" (Liu, 2003, Liu *et al*., 2004, 2005) which has attracted a great deal of attention from the scientific community worldwide, in part due to the simplicity and transparency of its criteria. In contrast with many rankings of international or national scope which have attracted a large audience (THES ranking, Ince, 2004), the Shanghai ranking is focused on the research dimension of universities and, therefore, relies heavily on "bibliometric" indicators. It also poses a provocative challenge to departments in universities or institutions in various countries that are involved in bibliometric indicators for institutional assessment, and which generally rely on a methodological background developed over decades by researchers in "scientometrics". Scientometrics lies at the crossroad of Informetrics, the Sociology and Economics of Science, and Science Policy Studies. It furnishes quantitative tools to the Social Sciences and to Science Policy, while providing indicators to decision-makers and stakeholders in Science and Technology. The implicit doctrine of scientometrics is that bibliometric indicators are coined in the substrate of Informetrics and Social Sciences. From their earliest publications, bibliometricians have issued warnings about the limits of indicators for evaluative purposes, while at the same time searching for ways to overcome them. A frank comment about the Shanghai ranking (from the point of view of professional scientometrics) is found in van Raan (2005). For these reasons, the Shanghai ranking presents a tempting occasion to examine some aspects of the state of the art of bibliometric indicators for purposes of evaluation and ranking.

It is useful at the outset to recall the premises of the Shanghai ranking. The exercise was meant to detect and to rank world-class institutional players, in order to position Shanghai Jiao Tong University and other Chinese universities in the international context, as well as (perhaps) to identify possible partners. The criteria and the weighting scheme are shown in Table 1.

Table 1 'Shanghai Ranking' indicators

Criterion	Indicator	Code	Weight
Quality of Education	Alumni of an institution winning Nobel Prizes and Fields Medals	Alumni	10%
Quality of Faculty	Staff of an institution winning Nobel Prizes and Fields Medals	Award	20%
	Highly cited researchers in 21 broad subject categories	HiCi	20%
Research Output	Articles published in *Nature* and *Science**	N&S	20%
	Articles in *Science Citation Index*-expanded and *Social Science Citation Index*	SCI	20%
Size of Institution**	Academic performance with respect to the size of an institution	Size	10%
Total			100%

* For institutions specialized in Humanities and Social Sciences such as London School of Economics, *N&S* is not considered, and the weight of *N&S* is relocated to other indicators.

** The total scores of the above five indicators divided by the number of full-time equivalent academic staff. If the number of academic staff for institutions of a country cannot be obtained, the total scores of the above five indicators is used.
For each indicator, the highest scoring institution is assigned a score of 100, and other institutions are calculated as a percentage of the top score. Scores for each indicator are weighted as shown to arrive at a final overall score for an institution. The highest scoring institution is assigned a score of 100, and other institutions are calculated as a percentage of the top score.

Source : Institute of Higher Education, Shanghai Jiao Tong University, 2004
http://ed.sjtu.edu.cn/ranking.htm

Since its appearance, a number of critiques of the Shanghai ranking have been expressed and were summarized at the first world-class university conference. Suggestions made included counting options for Nobel Prizes and Fields Medals and possibly including other scientific awards (an issue on the edge of bibliometrics). The question of the field-bias of the journals *Nature* and *Science* was also raised, since it obviously altered the relevance of this particular criterion. The significance and robustness of some indicators were called into question. The issue of institutional identification in the *International Science Indicators* (*ISI*) databases was another widely recognized critical issue in “ranking exercises”.

The first section will deal with some of these questions as well as issues that arise when the *ISI* database is used as a standard reference. The second section will focus on a couple of key questions in assessment studies, *i.e.,* comparability and scale. Scale issues concern the reference sets used for comparison as well as the size of the players themselves. In particular, it deals with a striking feature of the Shanghai ranking methodology: the emphasis it puts on *size-dependent* indicators. In the Shanghai ranking, *excellence* measures, which account for 70 percent of the weight, are in fact defined by the *number* of items (articles, authors, etc.) achieving a particular level. The next criterion, presence in *Scientific Citation Index* or *Social Scientific Citation Index (SCI/SSCI)* (20 percent of the weight), not an excellence measure, is also assessed in terms of *number* of items. Thus, 90 percent of the assessed weight is composed of size-dependent criteria.

In this section examination of the dependence of indicators on players' size will be made, and then measures of their dependence on thematic references (such as fields and scale-of-fields) will be taken. Some responses from a recent French assessment experience, "the Cooperative", will also be mentioned in this regard.

2 - Players' definition and identification - database issues

2.1. Who are the players ?

The question is twofold:
- which type and level of players are to be compared?
- once these have been determined, are the players correctly identified in the bibliometric database?

The decisions made in bibliometric studies in choosing a type or level of actor, for example the university level, has important consequences for interpretation. Are these entities well defined, corresponding to the same "player" in the international landscape of science? Some denominations such as university, department, laboratory seem universal and correspond to entities with some degree of autonomous governance. However, these organizational units vary widely in size, activity, and management characteristics, as well as across countries and/or disciplines. "University" is indisputably a legitimate level of comparison, but nevertheless heterogeneity of players and skewed distributions of size make accurate comparisons difficult (see section 3).

The second question concerns the identification of these players in bibliometric sources such as *ISI* databases, a problem that can turn into a bibliometrician's nightmare. The issue has two facets. First, the way in which authors describe their institutional affiliations in articles and how databases capture these affiliations, generally with little standardization or unification of institutional names, has an impact on bibliometric data. The same lab may appear under dozens of forms of its name. Secondly, one must take into account how these affiliations connect with the scheme of the survey. These questions are widely commented in the literature, and numerous case studies offer solutions for overcoming such difficulties. The problem is not equally serious in all research systems. Some countries, such as France, are well known for their particularly complex institutional setting. There, Public Research Organizations (PROs) are particularly important along with universities. They overlap in many hybrid forms *i.e.*, most French laboratories are joint laboratories, affiliated with several different institutions, PROs and/or universities. In addition, names, structures and affiliations of the laboratories change frequently.

All this makes it difficult to identify "from the outside" the articles produced by a particular research organization. It is nearly impossible to label correctly the production of various French research organizations without their active participation (through self-identification or confirmation of the identification). Among others, this is the case of the major French PRO, the CNRS (Centre National de la Recherche Scientifique). The indicators unit UNIPS (now IPAM) within CNRS established that 30 percent of CNRS publications are not labeled as such in the *ISI* database, and the trend is getting worse [i]. For this reason, the Observatoire des Sciences et des Techniques (OST), when constructing bibliometric indicators in the context of "the Cooperative", has called upon the research organizations themselves for assistance in the aim of building truly comparative indicators among research organizations (Esterle, 2005). Even if France represents an extreme case, other research systems have their own particularities. Thus, there is a sharp contrast between large-scale studies such as the Shanghai ranking, which cannot afford to undertake detailed, arduous player identification, and national assessment exercises where an accurate identification of players is necessary.

2.2. Dependence on databases

In most benchmarking studies performed to date, *ISI* databases have been used for building international indicators. They are excellent products in many respects. Their characteristics and possible "biases" have prompted a large number of studies beginning with the exchanges between Moravcsik (Moravcsik, 1985) and Garfield. The essential qualities of the base for mainstream research in Natural Sciences are rarely questioned. This being said, journal coverage choices made by *ISI* database personnel have their own limitations:

- The "tail" of *ISI* databases houses a heterogeneous population of journals with low visibility and high degree of national-orientation, which may jeopardize classical indicators of publication, citation, and impact. Ruling out tails on the criterion of journal language (van Leeuwen *et al.*, 2001), or more generally on criteria of journal impact and internationalization (Zitt, Ramanana-Rahary, and Bassecoulard, 2003) may be considered. In certain non-mainstream countries (*e.g*. Russia), measures of publication or impact can vary in opposite directions by a factor of two when journal sets are corrected. The way *ISI* manages the turnover of national journals also matters for emerging countries. Since half of the Shanghai ranking criteria make use of *ISI* data and the ranking is focused on "excellence", hence on the upper tail of distributions, the problem of delineation is not as critical as in other studies. However, the count for the *SCI* criterion is likely to be affected, especially for universities from non-mainstream or emerging countries, such as China and India.

- Mechanisms that might ensure a proper balance among fields are found in citation transactions. When transactions within or across disciplines become too small, especially for isolated fields or the low-impact tail, the interdisciplinary balance cannot be grounded on bibliometric considerations. This point represents a further limitation on studies that do not include disciplinary disaggregation.

- Extensions of *SCI* to Social and Human Sciences (*SSCI* and *A&HCI*) do not exhibit the same representativeness as *SCI*. Major reasons are the specificity of modes of production in Social and Human Sciences and the weight of national traditions (Hicks, 2004), as well as the fact that Social Sciences in non-mainstream or non English speaking countries may follow with a few decades' lag time the path of transition towards the international model[ii]. The situation is much different across disciplines.

Therefore, the choice of the *ISI* database for the ranking, although indisputable, calls for a cautious approach to standard delineation.

3. Dependence on size and scale

3.1. Players' size and variety of activity

By and large, bibliometric benchmarking deals with two types of indicators: size-dependent measures and size-independent (or at least primarily-independent) measures.

Size-dependent (SD) measures

Some bibliometric indicators, directly reflect – or are dependent on – the size of the player. In statistical terms, their expected value is an increasing function of the number of researchers, if one takes this as a convenient proxy for actors' size. Among SD indicators:

- the number of publications, which primarily depends on the amount of human resources.
- the number of citations, which depends primarily on the number of publications, itself linked to size
- the number of publications in some class of citations (e.g. the most cited decile in *SCI*)

In the Shanghai ranking, all criteria but the last one show a direct dependence on some aspect of size: number of graduating students for the criterion *"Alumni"*, number of staff (and financial resources) for "*Award"*, number of staff through number of publications for "*HiCi"*, "*N&S*", and "*SCI"*. Moreover, the last criterion termed "*Size"*, which introduces a "productivity" measure – not directly size-dependent – has been given a low weight.

A feature of SD measures is their expected co-linearity due to size. The authors of the Shanghai ranking also provide a direct correlation table (Table 2).

Table 2. Direct correlation among indicators

correlation	total score	*Alumni*	*Award*	*HiCi*	*N&S*	*SCI*	***Size***
total score	1.00						
Alumni	0.80	1.00					
Award	0.84	0.76	1.00				
HiCi	0.90	0.60	0.65	1.00			
N&S	0.93	0.67	0.70	0.86	1.00		
SCI	0.81	0.55	0.50	0.68	0.74	1.00	
Size	***0.83***	***0.68***	***0.73***	***0.70***	***0.77***	***0.56***	***1.00***

Sources: Liu, N.C., Cheng, Y.
data source: http://ed.sjtu.edu.cn/ranking.htm

As expected, high correlations are found among strictly bibliometric indicators on one hand ("*HiCi*", "*N&S*", "*SCI*"), and between "*Alumni*" and "*Award*" on the other. High correlations are observed between partial (esp. *"HiCi"*, "*N&S*") and total scores.

The high level of correlation among partial rankings builds a spurious robustness for the global ranking, but it is a built-in artefact. Adding or removing a SD criterion has little effect on a global ranking that is largely size-dependent.

'Size-(primarily)-independent'(SPI) measures

In some indicators, size is partially neutralized, usually by a size factor in the denominator of a ratio. These "size-(primarily)-independent" measures include the following:

- the bibliometric impact, by definition a ratio of citations to publications;
- the proportion of publications in some excellence class (examples: top cited class; *Nature* and *Science*; "leading edge");
- the scientific productivity measures, for specific input measures, again in ratio form. The indicator "*Size*", the only SPI measure used in the Shanghai ranking, belongs to this category.
- another type of institution productivity measure is any sensible central value of distributions of individual-level measures such as h-index (Hirsch, 2005).

The term "primarily" is important. In these ratio forms, only the linear part of dependence on the factor chosen is neutralized, while in econometric models of the input-output relation, or in scientometric models of the publication-citation relation in research systems (Katz, 1999) power-laws are dominant, leading to log-transformed models. This discussion forms part of the general debate about critical mass and increasing returns to size in science. In cases where increasing returns occur, SPI measures still convey a secondary effect of size, and the size effect carried by SD measures is magnified. However, the correlation production-productivity is far from being evidenced at all levels (for a recent study at the laboratory level (Bonaccorsi and Daraio, 2005)), and in any case it remains a statistical relation, with only a part of variance explained. In the Shanghai ranking, the correlation between productivity and SD measures is higher for excellence measures than for total publications: 0.50 for "*SCI*" – "*Size*".

How resistible is the size effect? If one looks at the *"HiCi"* criterion assuming, for simplicity's sake, constant returns, it turns out that if university "B" is half as large as university "A" then, all things being equal (same productivity in terms of articles/researcher), it will need *twice as many citations* in order to compete. In other words, it will need an impact figure twice as high as "A", which is quite a marked difference in terms of visibility. The mechanical advantage accruing to big players in the Shanghai ranking would be extremely difficult to compensate for by small entities, even with outstanding SPI performances.

How are size-independent measures operationalized? There is a sharp contrast between bibliometric impact and productivity measures. The former have been easily available from *ISI* citation indexes for decades, even though their technicalities and interpretation have given rise to a huge literature (reviewed by Glaenzel and Moed, 2002). This is not the case for productivity measures, which typically are difficult to establish especially at the macro-level. For measures of productivity based on publications, bibliometricians have issued some warnings about the numerator (see above the issue of delineation), but the problem lies more with the denominator, despite international guidelines (Frascati Manual and updates (OECD, 2003)). Barré gives some examples of difficulties of input measures (Barré, 2001). The question of data quality remains crucial while robustness of methods increases, for example DEA (Data Envelopment Analysis) techniques which allow composite outputs and inputs.

Growth: SD or SPI indicator?

If the heavy stress that has been placed on static aspects of size is questionable, the dynamic version of the indicators, that is, the *growth rate* of university resources and outputs, could be a more appealing measure. There is increasing attention being paid in the literature to growth dynamics of universities. The similarity of growth mechanisms in industrial firms and research structures is suggested by empirical models. There is some empirical evidence that growth rates and size are independent, following Gibrat's law, though size commands the width of the growth rates' distribution (Plerou *et al.*, 1999, Matia *et al.*, 2005). Thus, growth indicators could be considered as size-independent.

Dependence on organizational breakdown

Rankings may be different depending on whether the unit of observation is the university or some smaller unit, *i.e.,* schools, departments, or laboratories. The result depends on statistical features of the organization type.

Taking the hypothesis that the difference in university size is due to the total number of units (e.g. labs) of similar size that comprise a university, then size-effects consubstantial to SD-indicators will appear at the university level rather than at the component level. However, for some SPI-indicators, adverse phenomena may take place, namely for indicators based on means, *e.g.* bibliometric impact. For example, in international comparisons the distribution of mean impact of players, all things being equal, will tend to exhibit a larger variance in countries where players are small, whatever the reason may be: national specificity in organization, institutional dynamics, or the particular level of analysis chosen for the country. Populations of small players may be favored in particular indicators such as the proportion of players above an absolute threshold of impact (a classical "excellence" measure).

This being said, a few remarks are nonetheless in order:

- the hypothesis of equal size of fundamental units is unrealistic. There is some empirical evidence that the size of entities whatever the level shows a skewed distribution, and that the size of units and universities are linked according to power-laws, a feature not limited to scientific organizations (Matia *et al.*, 2005). Size effects are likely to be observed both at the university level and at the unit level;

- some degree of correlation is expected among the performances of units belonging to the same university as soon as the university successfully implements a policy of positioning in the international hierarchy of players. Organizational auto-correlations may be quite similar to spatial auto-correlations in spatial studies;

- further spatial aggregation (country-level) raises similar questions. In the Shanghai ranking scheme, the representation of countries is heavily dependent on national differences in size-distribution;

- some aspects of organizational breakdown issues can reflect – or compete with – "collective representations" of thematic breakdown as they appear in database classification or bibliometric mapping of themes (see below).

As quite different pictures may be obtained for various levels of organization (for example departments/schools and universities), one should remain very prudent when using a single-level approach.

3.2. Dependence on Field Delineation / Thematic Breakdown

Reference sets: from disciplines to research fronts

Bibliometric performance varies widely according to discipline. The "production function" of a publication differs greatly between say Fundamental Biology and Mathematics. The productivity by author is thus different across disciplines. In the same line, the average number of citations by publication is also quite different. A great deal of work accomplished in scientometrics since the 1970s by a number of research teams (ISI, CHI Research, ISSRU Budapest; CWTS, etc.) have highlighted these discrepancies, and suggested ways to keep things comparable, especially through field-normalization of indicators (*e.g.*, Schubert and Braun, 1986, 1996).

Indicators calculated to show the overall production of an organization, covering several disciplines, should be handled with caution. Such indicators are not well suited for ranking research organizations, especially those whose disciplinary profile is rich, unless they are properly normalized. As it turns out there is no way to avoid performing a discipline-by-discipline comparison, which then may serve as a basis for a normalized aggregate indicator that will be better suited for overall comparisons. Bibliometric bureaus have implemented a variety of normalized impact indicators, both on cardinal and rank approaches.

The Shanghai ranking criterion "*HiCi"* uses an *ISI* rank approach on a 21-level breakdown. In contrast, only global publication values are given for "*SCI*" criterion, without disciplinary count. A clear disciplinary bias, mentioned above, is carried by the selection of *Nature* and *Science*, which calls for a correction since players not active in topics privileged by these journals are under-represented.

The issue of grouping data by discipline or thematic area inevitably confronts bibliometrics with some difficult questions, as bibliometricians strive to take into account opposing criteria (organizational policy, scientific networks, methodological communities, etc.). For example, in a first-level exercise of slicing data into ten or so disciplinary fields, where does Computing Science belong? With Mathematics, Engineering Science, or set off by itself, or divided up for distribution between these two disciplines? The choice will deeply affect the picture of a country in the international landscape of science.

Here again there is no ideal solution, but rather choices to be made which may well have an exaggerated effect on the rank of a particular research organization depending on its disciplinary profile.

Local vs. global references

The set of journal where an actor publishes, or could publish – for example, by looking at a set of neighboring journals to publication journals – is a usual starting point. The obvious advantage of this approach is the precision it brings. Its main shortcoming is hypersensitivity to the concentration curve of the player, and the asymmetry of players' pair comparisons[iii]. The method is more adapted to benchmarking exercises, for the assessment of a few individual players.

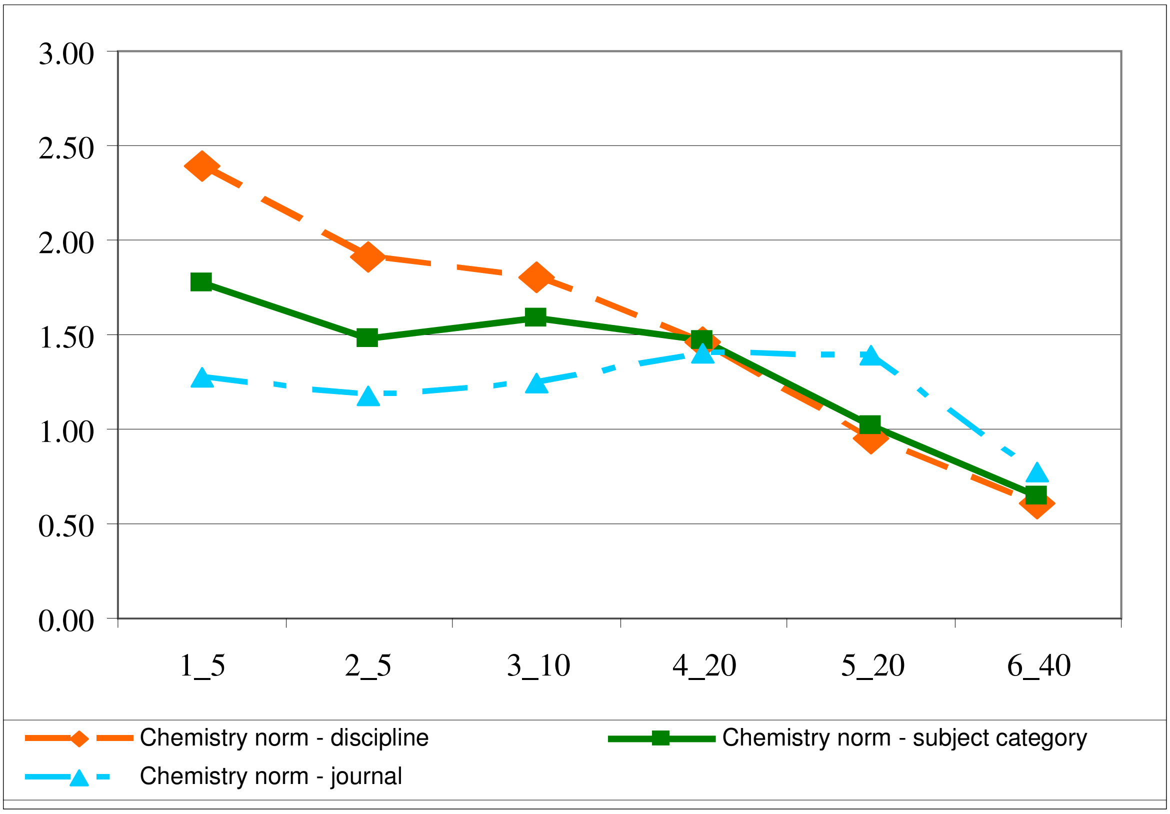
Another example of local reference is the design of thematic grids based on the organizational structures of particular countries or agents. Though being informative for particular benchmarking studies, the lack of generality of such approaches is penalizing. It is fair to recall that the field classification of "global" databases may bear some marks of their national origin, for example the *A&HCI*.

The choice of reference sets for normalisation

A comparable problem is that of the level chosen for developing a comparison. Should the set of references used for comparison be very narrow – micro-thematic areas (Kostoff, 2002) or journals, in a rationale of "relative-citation-ratio" indicators (Schubert and Braun, 1986, 1996) – in order to do justice to the variety of research profiles of various organizations? This option possibly grants an advantage to small organizations active in applied fields. Alternatively, should the comparison be carried out at an intermediate level? *ISI* "subject categories" are easily available and are often used by analysts. Or else, should it be carried out on a still larger frame of reference (sub-discipline, discipline) in order to take into account cross-disciplinary or generic research? This option gives an advantage to large organizations active in "big science". The question has been re-explored recently in a general way by OST (Zitt, Ramanana-Rahary, Bassecoulard, 2005) and several levels of normalization are now applied in OST analyses on universities and PROs. Figure 1 shows the citation profile of a university with three levels of normalization. In contrast with the "*HiCi*" ranking, relative measures of SPI-type are considered, namely the relative activity of the actor in classes of citations.

In the Shanghai ranking, an *ISI* 21-level discipline grid is used for the criterion *"HiCi"*, a relatively large-base normalization, which carries an implicit advantage for players active in basic and transversal research.

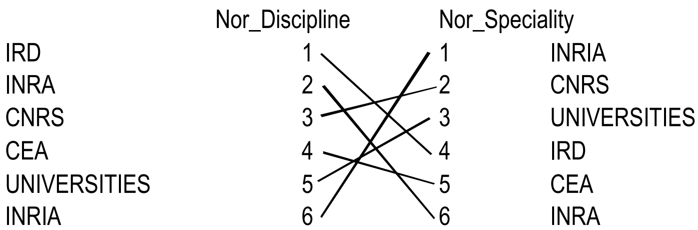
Fig.1 Activity index in citation classes: university X in chemistry - three levels of normalisation (discipline, subject category, journal).



Visibility classes (abscissas) are ranked by decreasing impact: 1_5 stands for the first excellence class in ISI source, with the five percent top cited articles, 2_5 the second excellence class with the next five percent, 3_10 the high visibility class with the next decile, etc. Ordinates represent the relative effort of the player in each class, as a percentage of the expected value (for example 1.5 if 7.5 percent of its articles are placed in the first five percent class). The more descending the profile, the more visible the player. Changing the level of normalization alters the profile of players. Here, the university X exhibits a fairly good profile at the journal level, but much better when the basis of normalization is enlarged. This means that the university is present in the most visible areas of Chemistry.
Source: OST commissioned study, unpublished

Changes of profile result in shifts in the ranking of players. For example, based on data on "the Cooperative", if one focuses on the excellence class and compares just two levels of normalization, the amount of change depends on the field. In fields with a homogeneous structure from the point of view of citing behavior (Physics for example), no ranking shift is noted, while ranking shifts are most evident (or "greatest") in Engineering Sciences (the latter displayed in Figure 2). Depending upon the structure of classifications (which may be macro-nomenclatures or bibliometric classifications), such changes may or may not have a significant effect on the ranking of players. There is no best way; one should accept several "zooming" levels on science.

Fig.2 Effect of scale on normalisation: French actors in top-rank impact class (engineering sciences)



Position shift in "excellence" class as measured by the ratio of each institutions/articles in the top class of impact. Left column: normalization at the 8-discipline level; right column: normalization at the specialty (="subject-category") level. The discrepancies are due to the heterogeneity of this field in citation structure amongst its subfields.
Source: Zitt, Bauin, Filliatreau, (2004).

It should be noted that the problem is even thornier when normalization is applied to emerging research organizations or to those publishing in emerging disciplines. A recent organization, for example, or one whose research is just getting started in a new direction will simply not show up in a ranking if the indicators employed are too generalist in nature. Here again, varying the focal point of the observation is critical.

The question might be extended to cross-field normalisation of a researcher's productivity.

Diversity

Bibliometrics and economics provide a number of indicators aimed at characterizing the production spectrum and the diversity of players' activities. Again the spectrum is defined according to some type of disciplinary or thematic reference, either being local to the university, or shared among players. For example, bibliometric breakdowns such as *ISI* subject categories, in the absence of international standards, often play the role of a reference. Obviously, measures of diversity are dependent on the type and the grain of these classifications.

The question also arises whether diversity is *as such* a performance. Whereas the quest for visibility and excellence can be held as universal, contrasted strategies of specialization exist, and the context and the scale do matter. For example, at the macro-level, the relatively specialized spectrum of mainstream countries (*e.g.*, the United States of America or the United Kingdom) in Biology is sometimes regarded as an evidence of leadership, while at another scale, the capability of differentiation of profiles in research or higher education supply is often viewed as a competitive advantage (Adams and Smith, 2003). The investigation of the complex relations between variety, size and growth will probably feed university studies in the coming years, as in industrial economics.

4 Conclusion: Big is (made) beautiful

The Shanghai ranking has provoked a great deal of interest, though some aspects of the methodology employed are in sharp contrast to conventional techniques of university assessment. A variety of criticisms by several authors have been leveled at the ranking on technical points, the most complete being probably van Raan (van Raan, 2005). Here the focus is on questions of size, scale and reference sets.

The Shanghai ranking favors “Size-Dependent” measures (90 percent of the total weight), a decision that mechanically favors big players, at the expense of smaller but high-performing players. If only “Size-Independent” ratios were used, various phenomena of returns-to-scale would already reintroduce a size-effect, but this time in relation with the usual acceptation of “academic performance”. The current weighting scheme “brings coals to Newcastle”. The authors of the Shanghai ranking do not ignore the problem. “*The Ranking Group is studying the possibility of providing separate rankings with and without the size indicator. For ranking with the size indicator, the weight of the size indicator could be as high as 50 percent. Furthermore, there are difficulties in defining academic staff and obtaining internationally comparable data*” (Liu and Cheng, 2005). One would wish to encourage Liu and Cheng in this direction, recalling that productivity indicators are not the only size-independent indicators and that bibliometric-relative indicators are quite easy to implement. Size-independent measures based on internal ratios can be built from the same data (putting aside the severe problems of data and institutional identification) for “*HiCi*” and “*N&S*”, or by adding estimates of human flows (“*Alumni*”, “*Award*”). “*SCI*” is hardly reducible without input data.

In summary, if *one* single ranking were to be used, it would be advisable to choose a different balance than the 90 percent SD-ten percent SPI ratio used by the Shanghai ranking group. Taking into account returns to scale and the related phenomenon of links between SPI and SD, this is almost entirely a size-driven ranking. A 50-50 mix ranking would still strongly reflect size, and even a 100 percent SPI ranking scheme would bear some imprint of size, due to returns to scale. Factor analysis could probably help to find independent combinations of criteria but the position and content of factors will evolve with time and make interpretations more difficult. Another solution is simply to admit two rankings, one reflecting SD only, the second based on SPI, expected to be correlated to a certain extent. The inclusion of growth rate, apparently not connected to size, would also be worth considering although this approach presents some methodological difficulties. Variety, whatever the measure, is an important characteristic of research systems and should be shown, but its relation to growth and performance indicators is so complex, and changing with scale, that it seems dangerous to consider it as a performance indicator on its own.

A few events, in scientometrics, have a feedback effect on scientific community to such an extent that its behavior changes markedly. The higher magnitude star is the "impact factor" by Gene Garfield, which affected the evaluation process in most institutions in the 1960s. The fact that the measure was slightly flawed (Moed, 2002) does not diminish its historical importance. There are other examples that show that the scientific community is adaptable and reacts to scientometrics-based evaluation systems, including when they are clearly sub-optimal, as shown in a recent example studied by Butler (2003). It is too early to assess the consequences of Shanghai ranking diffusion, but its wide diffusion is likely to trigger adaptation reactions as well. If the ranking keeps its emphasis on size, some players might try to grow – and merger may be the quicker way to do it – or to look bigger, as by some re-labeling policy. It might result in favorable outcomes. However, it would be more efficient if players exhibited an adaptive behavior based on improved versions of the academic ranking process.

ACKNOWLEDGEMENTS

The authors wish to thank Elise Bassecoulard and Suzy Ramanana-Rahary for their help.

REFERENCES

ADAMS J., SMITH D. (2003), *Funding research Diversity*, Universities UK, Evidence Ltd, London (UK).

BARRÉ R. (2001), Sense and nonsense of S&T productivity indicators, *Science and Public Policy*, 28 (4): 259-266.

BONACCORSI A., DARAIO C. (2003), A robust nonparametric approach to the analysis of scientific productivity, *Research Evaluation*, 12 (1): 47-69.

BONACCORSI A., DARAIO C. (2005), Exploring size and agglomeration effects on public research productivity, *Scientometrics*, 63 (1): 87-120.

BUTLER L. (2003), Explaining Australia's increased share of ISI publications - the effects of a funding formula based on publication counts, *Research Policy*, 32 (1): 143-155.

ESTERLE L. (2005), Comparing and Evaluating Public Research Organisations: A unique, participatory mechanism in place in France, *Research Evaluation*, 14 (2): 129-135.

GLAENZEL W., MOED H. F. (2002), Journal impact measures in bibliometric research, *Scientometrics*, 53 (2): 171-193.

HICKS D. (2004), The four literatures of Social Science, In: W. GLAENZEL, H. MOED , U. SCHMOCH (Eds), *Handbook of Quantitative Science and Technology Research*, Kluwer Academic Publishers, pp. 473-496.

HIRSCH J.E. (2005), An index to quantify an individual's scientific research output, PNAS, 102(46),. 16569-16572.

INCE, M. Coord. (2004), *World Universities Ranking*, The TIMES, Higher Education Supplement (THES) Nov. 5

KATZ J. S. (1999), The self-similar science system, *Research Policy*, 28 (5): 501-517.

KOSTOFF R. N. (2002), Citation analysis of research performer quality, *Scientometrics*, 53 (1): 49-71.

LIU N. C. (2003). Studies of world-class universities in the Institute of Higher Education of Shanghai Jiao Tong University, *Tsinghua Journal of Education (China)* 24 (3) (in Chinese).

LIU N. C. et al. (2004), *Academic ranking of world universities*, Institute of Higher Education, Shanghai Jiao Tong University, Top 500 World Universities. http://ed.sjtu.edu.cn/ranking.htm (last visited 2005/08/12)

LIU N. C., CHENG Y. (2005) Academic ranking of world universities - methodologies and problems, *1st Intl Conf. on World-Class Universities (WCU-1)*, Shanghai (CHN).

MATIA K., AMARAL L. A. N., LUWEL M., MOED H. F., STANLEY H. E. (2005), Scaling phenomena in the growth dynamics of scientific output, *Journal of the American Society for Information Science and Technology*, 56 (9): 893-902.

MOED H. F. (2002), The impact-factor debate: the ISI's uses and limits, *Nature*, 415 (14 Sep): 731-732.

MORAVCSIK M. J. (1985), *Strengthening the coverage of Third World Science. The final report of the Philadelphia Workshop and of the discussions preceding and following that workshop*, Institute of Theoretical Science, University of Oregon, Final Report.

OECD (2003), *Frascati Manual 2002 - Proposed Standard Practice for Surveys on Research and Experimental Development*, OECD Publishing.

PLEROU V., AMARAL L. A. N., GOPIKRISHNAN P., MEYER M., STANLEY H. E. (1999), Similarities between the growth dynamics of university research and of competitive economic activities, *Nature*, 400 (6743): 433-437.

SCHUBERT A., BRAUN T. (1986), Relative indicators and relational charts for comparative assessment of publication output and citation impact, *Scientometrics*, 9 (5-6): 281-291.

SCHUBERT A., BRAUN T. (1996), Cross-field normalization of scientometric indicators, *Scientometrics*, 36 (3): 311-324.

VAN LEEUWEN T. N., MOED H. F., TIJSSEN R. J. W., VISSER M. S., VAN RAAN A. F. J. (2001), Language biases in the coverage of the Science Citation Index and its consequences for international comparisons of national research performance, *Scientometrics*, 51 (1): 335-346.

VAN RAAN A.F.J. (2005), Fatal attraction: Conceptual and methodological problems in the ranking of universities by bibliometric methods, *Scientometrics*, 62 (1): 133-143.

ZITT M. (2005) Facing diversity of science: a challenge for bibliometric indicators - comments on A. Van Raan's focus article, *Measurement: Interdisciplinary Research and Perspectives, 3,1, pp. 38-49*

ZITT M., BAUIN S., FILLIATREAU G. (Coord.) (2002, 2004). *Bibliometric indicators of public research institutions* (in French) ; Indicateurs bibliométriques des institutions publiques de recherche. http://www.obs-ost.fr.

ZITT M., PERROT, F. BARRE, R., (1998) The Transition from "National" to "Transnational" model and Related Measures of Countries' Performance, *Journal of the American Society for Information Science*, 49, 1, 30-42.

ZITT M., RAMANANA-RAHARY S., BASSECOULARD E. (2003), Correcting glasses help fair comparisons in international science landscape: Country indicators as a function of ISI database delineation, *Scientometrics*, 56 (2): 259-282.

ZITT M., RAMANANA-RAHARY S., BASSECOULARD E. (2005), Relativity of citation performance and excellence measures: From cross-field to cross-scale effects of field-normalisation, *Scientometrics*, 63 (2): 373-401.

---

[i] Bauin S., pers. comm.

[ii] For natural sciences, the transition phenomenon is rather important for interpreting long-term series of output indicators (see, for example, Zitt, Perrot & Barré, 1998)

[iii] "A" and "B" are not rated in the same way in the journal set local to "A" and in the journal set local to "B".