University rankings are a relatively new phenomenon in higher education. Although quite an established practice in the U.S., it is only within the last decade that attempts to analyse university performance have spread to the rest of the world, and that we also have seen new global rankings appear – rankings attempting to measure university performance beyond national borders. No wonder that this trend is accompanied by a growing interest in studying rankings throughout the world. This book is written as part of the effort to better understand rankings and their effects on higher education.

A serious approach towards university rankings implies that rankings should be analysed properly, including the methods used and the indicators chosen, and investigate the objectives claimed. If university rankings are considered as consumer information then everyone should have an interest in basing such guidance on valid and reliable data and methodology. A serious analysis should also discuss the wider implications of rankings as an emerging phenomenon in higher education.

Consequently, the contributions to this book investigate and analyse how different rankings work, how they reach their conclusions, and on what data and methodology they are built. Furthermore it provides a critical reflection about the impact of rankings on higher education, how and in what way rankings influence policy-making, the structure of the sector, or the internal life of the sector.

This volume provides insights in university rankings for stakeholders in higher education and policy as well as for researchers, students, and journalists.
University Rankings, Diversity, and the New Landscape of Higher Education
Global Perspectives on Higher Education provides cogent analysis and comparative perspectives on these and other central issues affecting postsecondary education worldwide.
University Rankings, Diversity, and the
New Landscape of Higher Education

Barbara M. Kehm
University of Kassel, Germany

Bjørn Stensaker
NIFU STEP, Norway
TABLE OF CONTENTS

Introduction (Bjørn Stensaker and Barbara M. Kehm) vii

PART I: METHODOLOGY

A Global Survey of University Rankings and League Tables 3  
*Alex Usher and Jon Medow*

Reputation Indicators in Rankings of Higher Education Institutions 19  
*Gero Federkeil*

World University Rankings. The Need for a New Paradigm 35  
*Roland Proulx*

Institutional Rankings, Marketing, and the Needs of Intending Students 47  
*Bernard Longden and Mantz Yorke*

Disseminating the Right Information to the Right Audience: Cultural Determinants in the Use (and Misuse) of Rankings 65  
*Leon Cremonini, Don Westerheijden, and Jürgen Enders*

PART II: IMPLICATIONS

Global University Rankings: Some Potentials 85  
*Simon Marginson*

Convergence and Diversity: The Role and Influence of University Rankings 97  
*David D. Dill*

The ‘World-Class’ University in Europe and East Asia: Dynamics and Consequences of Global Higher Education Reform 117  
*Rosemary Deem, Lisa Lucas, and Ka Ho Mok*
TABLE OF CONTENTS

Towards Unified and Stratified Systems of Higher Education? System Convergence and Organizational Stratified Differentiation in Europe 135
Massimiliano Vaira

Between Over-Diversification and Over-Homogenization: Five Decades of Search for a Creative Fabric of Higher Education 155
Ulrich Teichler

About the Authors 183
INTRODUCTION

WHY A BOOK ABOUT RANKING?

Those interested in higher education worldwide have noticed the rather frequent way universities nowadays are referred to in TV, newspapers or on the internet. Through numerous rankings, guides and report cards, we are now learning which university is simply the best; in teaching and/or research, in a particular country, or within a region. Those universities that are not the best – and they are many – do not seem to be very pleased when journalists ask them why they are not on the top. They reply that the published rankings are unfair, methodologically unsound and measure only a small fraction of what universities do. Of course, if one of these universities actually ends up in a comfortable position in one of the many rankings around, they would perhaps also claim that this accomplishment is well-deserved, and mirror the good work that has been done over a number of years.

Hence, to most of the public, and perhaps also to many academics working in universities, the emergent rankings are probably considered as nothing more than an entertaining read. However, those that claim to be interested in the results of the existing rankings would probably argue that such rankings should be considered as an important development and much-needed consumer information about which universities provide the best quality education – information that should be relevant for the choices students make about where to study, information that should be the basis for decisions as to which university should receive public money for future research, or which university provides most value for money. As a key institution in society one could argue that universities, as all other service providers, should be accountable for what they do and what they accomplish.

This book does not question the legitimate need for universities to report to the public what they do, and to provide evidence about the results they achieve – in short, they are confronted with the question: how do you know that you do what you say you are doing? On the contrary, this book is written because of the need to take the public accountability functions of universities seriously. However, an approach of such a serious nature implies that rankings should be analysed properly, including the methods used and the indicators chosen, and investigate the objectives claimed. If university rankings are considered as consumer information, then everyone should have an interest in basing such guidance on valid and reliable data and methodology. A serious analysis should also discuss the wider implications of rankings as an emerging phenomenon in higher education. Are rankings just consumer information, or can they also affect higher education in other ways? And more importantly, are we then talking about positive or negative consequences?
There is growing interest in studying rankings throughout the world (see also special issues of *Higher Education in Europe* in 2005 and 2007, Sadlak and Liu 2007). This book is written as part of the effort to understand rankings better, and as a response to the many unknown answers to the questions that have been raised earlier.

THE RATIONALE FOR RANKINGS AS A GLOBAL PHENOMENON

Rankings, university guides and various report cards are a relatively new phenomenon in higher education. Although one can argue that university guides are quite an established practice in the United States of America, it is only within the last decade that similar attempts to analyse university performance have spread to the rest of the world, and that we also have seen new global rankings appear – rankings attempting to measure university performance beyond national borders.

The simple answer to why we have witnessed this development is the aspiration to make money. Most rankings and university guides are published by commercial media companies, which see this as a new way to make a profit and increase readership. A more complex answer would emphasise at least three other major factors. First, rankings may be said to be a response to the expanding sector of higher education. Higher education is one of the fastest growing sectors globally, with more and more students, especially in Asia, entering universities. This almost global transition from elite systems to systems of mass and universal higher education has also increased the market potential of the sector itself. And in higher education systems in which the decision to enter a university represents a considerable private investment, for example in the form of tuition fees, the student group is now so large that commercial media companies view them as an interesting market segment in society. Whether the majority of students actually use rankings for selecting the institution where they want to study is, in this respect, not the most important issue.

A second and related explanation would emphasise the fact that the providers of higher education are also becoming more numerous. Following the demand for higher education, new institutions of higher education are established almost every day around the world. Since these are newly established institutions without a history, without academic traditions and without a respectable reputation, one could argue that rankings would have the rationale of producing valuable information about the quality and performance of these institutions. Whether rankings actually focus on these institutions is again perhaps not the most important issue.

Third, in parallel to the expansion of higher education, not only in terms of people and institutions, but also in terms of resources spent, the sector has, in numerous countries, experienced that those funding the activity (in most cases national governments) have launched a number of reforms to make the sector more efficient and effective. For many higher education institutions, this has meant more autonomy followed by more responsibility for accomplishing the objectives set for the sector. Such reforms can be said to have paved the way for the rankings. In some countries the reforms have changed the ways in which higher education is perceived, for example, by challenging the conception that all public higher education institutions
are of equal quality, or by establishing evaluation processes, performance indicator systems, and reporting schemes providing those that create the rankings with most of the information they need to make their comparisons. However, whether the information used to feed the rankings is meaningful in terms of serving as indicators of performance and quality could be contested.

As indicated, the rationales behind the establishment of rankings are not only linked to extrinsic factors outside higher education, but also to intrinsic factors within the sector. Hence, in principle one could argue that rankings are a sign of a radical transformation of higher education as such. Of course, how one should interpret this transformation, and the place rankings have as part of the transformative process can be disputed.

THE DIFFERENT INTERPRETATIONS OF UNIVERSITY RANKINGS

There are at least five different interpretations of what rankings mean for higher education, and how rankings might contribute to change the sector. These five perspectives are not mutually exclusive, but are presented separately to illustrate the many potential functions of rankings in higher education.

Rankings as Market Regulation

One interpretation of what rankings mean for higher education is associated with the changes in how the sector is regulated and governed. In this perspective, the state has abdicated as the key regulator of higher education, and the governance of the sector is more diffuse and complex than in the past, creating hybrid governance arrangements (Gornitzka and Maassen 2000). As part of these hybrid governance arrangements, the state has facilitated a number of market-based or quasi market-based governing strategies, including increased competition between higher education institutions. Rankings can be said to be linked to this governing approach in a number of ways. An interesting aspect here is the fact that some countries link institutional positions in ranking exercises to resource allocation. The Research Assessment Exercise (RAE) in the United Kingdom, and the Performance-Based Research Fund (PBRF) in New Zealand are prime examples (Clarke 2005), a development that can be interpreted to represent an ‘outsourcing’ of the instruments of governance. While such a direct link between ranking exercises and governmental funding is still rather unusual, it is intriguing since the governments in such cases also may pave the way to align their policy-making to the design of (global) rankings. Even more intriguing are the potential consequences of this transfer, if one also considers the possibility that some of the rankings might be based on questionable methodology.

The role of rankings in this perspective is not only that of being a means in a more market-regulated sector, where rankings could be considered an instrument for regulation. One could even argue that rankings may also be a driver of the development. Since rankings are open, available and easy accessible to the public, the broader public also gets more insight into the activities and into the performance
of higher education. A likely consequence is that the public may more easily develop its own opinion about the quality of the sector, and that they subsequently want, or will be given a stronger voice in the policy-making process (see also Alderman and Brown 2005).

RANKINGS AS GLOBALISATION

An alternative perspective would see rankings as part of the increasing process of globalisation of higher education. Although most rankings are national in focus, most attention is undoubtedly given to those rankings that have a global focus. This perspective has less emphasis on the abdication of the state. The focus is instead on the emerging global economy, trade liberalisation and increased mobility among students and academic staff. Within this global economy, higher education is regarded as just another economic sector (Gumport 2000), of which it is important to collect, systematise and compare the service providers. From being a mainly nationally oriented sector, most higher education institutions nowadays have a strategy for internationalisation, and taking part in that also implies the need for greater transparency (Huisman and van der Wende 2004).

The role of rankings in this perspective is that of offering consumer information to the many buyers and sellers participating in the globalised game of higher education services. Global rankings offer players of the game information which claims to compare higher education institutions across national borders. As such, rankings are part of the standardisation of institutions in society (Brunsson et al. 2000) with potentially negative consequences for institutional diversity. The difference to the former perspective is that while the state in a market perspective allows market mechanisms to play a role, it is the market that is ‘in charge’ in the latter perspective, restraining the possibilities of the state to interfere in the governance of the sector.

RANKINGS AS THE RISE OF THE AUDIT SOCIETY

Given the expansion of higher education in most countries in terms of the number of students that enter university education and the resources spent on the sector, an increasing interest has emerged in the 1990s as to whether the sector delivers with respect to outcomes and results. While discussions about effectiveness and efficiency in the past often could be characterised as a closed conversation between the state and the universities, this discussion has been opened up as a consequence of the rise of the audit society (Power 1997). A key factor leading to this development is again the expansion of higher education, and not the least, due to the many new providers entering the sector. As in many sectors in society, quantitative expansion often triggers worries about the quality achieved.

In the last decades, higher education has been confronted with a number of methods and schemes aiming at assessing quality (Westerheijden et al. 2007). However, although we have witnessed a considerable spread of quality assurance schemes in
this period, one could still argue that quality assurance has failed in addressing issues concerning learning outcomes, performance and results (Stensaker 2003).

In an audit society perspective, rankings then could be seen as a new accountability mechanism in higher education – a mechanism that has developed because existing mechanisms have not been able to answer the critical questions of the public. While the accountability function of rankings could be interpreted within a market perspective, one can nevertheless also argue that rankings fulfil an important democratic function in society – as an independent source of information about a sector that is becoming more important.

**Rankings as Institutional Identity Creation**

If we change the level of analysis from a macro level to the institutional level, quite a different perspective on rankings emerges. While many rankings proclaim their purpose to be that of affecting student choice and student behaviour, one should not rule out the possibility that those mostly taking action vis-à-vis rankings are the higher education institutions themselves – in a race for prestige and position in the academic pecking order (Dill and Soo 2005). Rankings provide institutions with prestige, depending on their position in the hierarchy created. As Labianca et al. (2001) have pointed out in a study of university emulation in the United States, universities consider other universities’ reputation and position as crucial for their own strategic development.

In this perspective, rankings function as a fashion arena in which institutions compare themselves, and in which they strive to create and build their own identity. As in other fashion arenas, rankings define who is hot and who is not, due to the emphasis on hierarchy and exclusivity (the haute couture). Whether the identity intended to be built is useful or relevant in a systems perspective is of less importance here. Fashion creates its own field with quite different norms, values and logics, than in other societal arenas.

**Rankings as a Symptom of the Knowledge Society**

While a globalisation perspective emphasises the quantitative expansion of higher education, the mobility of knowledge and trading within the existing knowledge base, a knowledge-society perspective would rather emphasise the qualitative changes in the production and structuring of knowledge. In the latter perspective, the innovative aspects of knowledge are valued and appreciated – how knowledge can be applied and exploited commercially in the market place (Gibbons et al. 1994). As knowledge producers, universities for centuries have provided ideas and theories relevant for many different purposes and contexts. Rankings may contribute to change this picture.

Hence, in this perspective ranking is a symptom of the emerging knowledge society, and the tendency to emphasise certain dimensions of the knowledge production. The function of ranking in this perspective is that of a structuring device of knowledge. Although the external information function is part of this
perspective, the important aspect is how higher education may change internally as a result of this development. Rankings, through their emphasis on certain indicators including research, patents, employability and the links between higher education and industry may contribute to change the conceptions about what kind of knowledge is considered as valuable, and how knowledge should be structured (Lyotard 1984). Somewhat paradoxically, the idea about the knowledge society may actually restrain our beliefs about the many values of knowledge with rankings playing the role as the normative filter.

THE EPISTEMIC BASIS FOR RANKINGS

Based on the many potential functions rankings may have in higher education, it is an important task to analyse the epistemic basis for existing rankings. While rankings often proclaim a common-sense approach to the primary processes of higher education, emphasising what is claimed to be objective and not self-reported data and giving priority to indicators with considerably legitimacy (for example, Nobel Prizes), a serious epistemological analysis should also question our belief system, our knowledge about the sector and our attitudes towards the methods and strategies used to create the rankings. Recently, initiatives have been taken to establish some sound standards as to how rankings should be carried out (see International Ranking Expert Group 2006), but the beliefs about higher education, on which many rankings are based, are perhaps the aspect most difficult to challenge. A starting point here could be the fact that prior to the appearance of any published ranking, there actually exist considerable tacit knowledge inside and outside of higher education, about which institutions could be considered to be the ‘best’ and the most ‘excellent’ – not only nationally, but also internationally. Beliefs that many rankings today replicate and seem to take for granted. Given the many faculties, schools, and departments inside a university, one could question the belief of whether every unit can be ‘excellent’ – still a considerable number of rankings do not break down data from the institutional level. The picture we are presented with then is an ‘imagined average’ – imagined, because few of the single units would fit to the overall position that an institution is given in the rankings.

Other beliefs which are important for the construction of the rankings are related to the relationships between research and education, between formal competence and student learning and to the importance given to certain dimensions of the knowledge production in universities. Such beliefs are frequently visualised through indicators such as student–teacher ratios, resources spent on staff and students or various indicators concerning the infrastructure. The links between these other related indicators, and the quality of the provision offered are, at best, rather uncertain. A related problem is that many indicators used in rankings could be characterised as input factors providing little information of the outcome, for example, in terms of student learning. Even indicators that seem to be output-based, such as research publishing, should actually be perceived as input indicators if they are used to establish a causal relationship between research and teaching.

A third problematic issue concerning rankings is related to the way indicators are selected, weighted and given priority in the overall analysis of the available
data. Many ranking use data collected for other purposes in their analyses, without checking such data for validity and reliability. Furthermore, although not all rankings prioritise among indicators, most of them do. However, these weightings seem to be a rather subjective process. Although a given reader is informed about the fact that certain indicators count more than others, the reason for the same and the technicalities associated with the actual weight given is, in most cases, not disclosed.

A final area for concern relating to the epistemic dimension of rankings is the assumption that knowledge can be deconstructed, quantified and summarised in overall scores. While writers on higher education for a long time have argued that the knowledge created in universities is contextual, integrated, and culturally embedded (see e.g. Clark 1983), rankings treat knowledge almost as an externality of the university.

THE STRUCTURE OF THE BOOK

Based on the introduction so far, it is perhaps not surprising that this book is divided into two parts. The first part takes the latter issue of the epistemic nature and basis of rankings one step further, investigating and analysing how different rankings work, how they reach their conclusions, and on what data and methodology they are built. The second part picks up our invitation to start a critical reflection about the impact of rankings on higher education, how and in what way rankings influence policy-making, the structure of the sector or the internal life of the sector.

A modern standard reference to ranking systems and their characteristics has been the survey done by Alex Usher and Massimo Savino in 2006, in which they conducted a large-scale comparison of a number of ranking systems all around the world. In this book, Alex Usher has joined up with Jon Medow to update their previous survey. Compared to their 2006 version, nine more rankings are added, and changes in methodology in some of the ones surveyed in 2006 are also taken into account. As such, the new survey provides an impressing overview of central characteristics of ranking systems, although a key message by the authors actually is that diversity rather than uniformity still dominates the different rankings. However, the puzzle for the authors is that despite this diversity, how is it that some familiar universities such as Harvard, Cambridge, Stanford and Oxford always dominate the top places of different rankings? Although they are not able to arrive at a decisive answer, they do point out that ‘age’, ‘faculty size’ and ‘per-student expenditure’, are factors that are likely suspects in explaining why some institutions always come on top. Following from this, the chapter concludes by arguing that the future of rankings probably lies in the growing amount of information from institutional sources, which may further increase the number and types of rankings available. Usher and Medow discuss whether this development to some extent may contribute to ‘democratise’ rankings in that more power is transferred to the consumer and the user of rankings.

Interestingly, one could argue that consumer or user power is already taken into account in some of the rankings published by the weight given to the views by academics and students on the quality of various higher education providers. Hence, in other sectors of society, targeting the consumers or users of a given service...
about their views of the services offered is in general regarded as the common-
sense approach. One could argue that with respect to university rankings reputation
is an obvious way to provide information about the sector. In his chapter, Gero
Federkeil still argues that there are quite a number of problems with using reputa-
tion indicators in a higher education setting. The most problematic aspect is
perhaps the tendency to use reputation as a proxy for quality. Federkeil offers an
explanation for the interest in reputation in higher education by linking the concept
to inherent characteristics of the sector, but points out that although reputation is a
symbolic aspect with much attention attached, it does not make the concept more
valid or reliable as an input to ranking exercises. In his chapter he shows how
reputation may differ significantly between departments within the same institu-
tion, between different groups of stakeholders and between and across countries.
The consequences of (mis-)using reputation indicators in rankings are further
elaborated, and it is shown how reputation indeed is a social construction, meaning
that it is impossible to find a joint agreement on reputation as such; one can only
come across an agreement on reputation among a particular sample of people in a
particular context.

In the following chapter, Roland Proulx picks up the issue about reputation
indicators in rankings, and points to the links between global university rankings
and knowledge economy in a global age as an explanation for the emphasis given
to reputation. To deal with the problem, Proulx suggests a reconstruction of global
rankings by including benchmarking into the ranking process to ensure that
comparisons made are comparing what is actually comparable. This would necessi-
tate comparison at the programme or basic unit level, (department/faculty) or at the
level of disciplines, rather than at institutional level. Proulx then goes on to discuss
the advantages of benchmarking exercises in terms of generating valid indicators,
choosing variables and definitions and validate the data and information generated.
The insistence on quantitative measures and benchmarking in the establishment of
global university rankings is seen by Proulx as a new paradigm which might serve
the purpose of rankings better, and provide the public with more reliable and valid
information.

Providing the public, most often students, with better information about the
quality of higher education providers is often the rationale behind the growth of
ranking systems. Quantitative and precise information about central characteristics
of universities and colleges may, in this respect be seen as the right way forward. However, as Bernard Longden and Mantz Yorke point out in their chapter, the
existence of certain quantitative data sets, and our trust and sense making of such
data can be problematic. While quantitative data in general are interpreted as more
‘trustworthy’ than qualitative data, Longden and Yorke show how changes in
university positions in a given ranking may depend on the way data are collected
and/or reported, and that the numbering of rankings may over-exaggerate what
in reality may be quite small differences and changes. Longden and Yorke also
acknowledge the claim that rankings are carried out to provide intending students
with better information on which they can base their choices in terms of institution
and programme, but they argue that given the way some quantitative data are
systematised and presented, the result may actually be potentially misleading to
students, and have serious negative implications for a given institution. Three conclusions are drawn from the analysis. First, more attention should be paid to the publication of what a university is able to offer to international students; second, there is a necessity for data management within an institution; and third, ranking statistics and data should be offered in a customised form, or in such a way that users may interact with the original data.

The issue concerning the relevance of rankings in an international setting is also a key concern for Leon Cremonini, Don Westerheijden and Jürgen Enders in the chapter that ends the methodological part of the book. Cremonini and his colleagues start out by claiming that rankings – or report cards as they prefer it to be labelled – often are designed according to a particular cultural context that may be of less benefit to students or people from other cultural contexts. Acknowledging the importance of rankings for university and college choice, the authors point out that report cards lack an important element which often determines choice, namely the issue of ‘culture’ and cultural differences. They argue that indicators used in report cards may be interpreted differently by different social and cultural groups, and that many rankings are currently dominated by the Western ‘cultural conventions’. The authors see the problems in handling this challenge, but do argue that new technology and the possibilities of ‘targeting’ information to particular groups could make report cards more relevant as an information source for a more multicultural society.

While the five chapters that mainly focus on methodological issues, validity and reliability of rankings provide a comprehensive and multifaceted description of central characteristics and shortcomings of modern rankings, the five chapters in the second part of the book address issues of the potential links between rankings – good or bad ones, irrespectively – and the future development of the sector in terms of diversity and structure. The point of departure for this section is the concern that rankings may be a powerful, although still not the only, driving force contributing to change central values and functions of higher education, and consequently also the whole organisation of the sector. While diversity – in missions, academic profiles and links to society – in general has been seen as an important aspect of higher education (Clark 1983), rankings may contribute to change this considerably. The argument often raised is that the emphasis on research and research output may create more similar institutions unable to cater to the diverse needs of a multicultural and complex society. However, as pointed out in the chapters in the second part of the book, there are also other potential implications of rankings that should be taken into account.

While the chapter by Simon Marginson also dwells on some of the methodological weaknesses of rankings, the focus of this chapter is first and foremost on their contributions to what Marginson calls the public or private good for the society at large. By using these terms as analytical tools, Marginson investigates and identifies those benefitting from rankings. He shows how rankings sometimes redistribute public and private goods in the higher education market, and how this may result in fewer public goods, but quite a number of private goods – especially for those universities which fare well in the rankings. One of the main problems,
according to Marginson, is that many rankings become an end in themselves with few or absent linkages to quality improvement of the sector as such.

The research emphasis that characterises many rankings are also a concern of David Dill when he suggests that a potential implication of current rankings is an exceedingly expensive ‘academic arms race’ for prestige which more and more substitutes the measurement of real performance. One important implication of such an arms race is increased costs for students (tuition fee) and/or a drive to increase the expenditures of higher education in general – once again reducing the public good. Dill argues that one runs the risk of developing universities that only become quite costly ‘screening’ mechanisms for the labour market rather than producers of human capital and new knowledge. Hence, the author concludes that government regulation may be needed to offer alternative sources of information to students and others who want to know what learning outcomes universities may provide.

Whether or not particular governmental regulations may have an impact on the rankings can still be questioned, since the rankings that are ‘going global’ seem to strengthen their position and importance. This is an issue that is discussed by Rosemary Deem, Lisa Lucas and Ka Ho Mok in the chapter in which they identify some of the challenges the new expectations related to the idea of ‘world-class universities’ may create for European and Asian higher education institutions. Although Deem and her colleagues do point out that the meaning of the concept ‘world-class’ is somewhat fluid and also depends on the indicators used, in both of these world regions, competitiveness and the strife for excellence and international standing has clearly increased. As one of the results, the authors note a concentration of research funding to ensure world-class status for a few universities to the detriment of the majority of institutions which have to struggle to maintain their level of funding. Such a negative consequence can endanger the (national) system of higher education as a whole because governments may neglect the losers in this race, and even the world-class institutions will have to rely on the strength and reputation of the system as a whole. For some – especially developing countries – the authors argue that global rankings may actually have damaging effects for the development of more comprehensive and domestically relevant higher education systems for the sake of international competition.

In the last two chapters of the book, rankings and their implications are contextualised and related to current developments of the sector, and how the landscape of higher education may look in the future. The development of higher education systems is the key concern of Massimiliano Vaira, who in his chapter takes a closer look at European higher education and the current restructuring process of this sector caused by the Bologna process, policy copying, and growing internationalisation. While some contributors to this book have pointed out the need for governmental action to be taken to counteract some of the unwanted consequences of rankings, Vaira argues that it is also possible to see how many ranking schemes fit well with governmental policies, emphasising more differentiation and stratification of the sector. In his chapter Vaira points out that rankings also influence national policy agendas both in terms of introducing policies for stronger differentiation, and of legitimising such initiatives politically. Hence, the argument launched is that the
INRODUCTION

influence of ranking systems on national policy making should not be underesti-
imated, and may cause ‘biased’ policy initiatives to be taken – not only in developing
countries as Deem as her colleagues argue – but also in well- developed countries.
Although Vaira is careful in his conclusion, the chapter demonstrates effectively
how ranking systems are intertwined with a number of other developments and
policy initiatives, not only affecting higher education institutions, but also the
policy framework governing the sector.

The final chapter, by Ulrich Teichler, relates rankings to larger historical devel-
opments of higher education, and points out that although ranking currently is
enjoying the centre stage of academic and political attention, the higher education
sector has proven to be quite robust and resistant to a number of extrinsic attempts
to instigate change over time. This does not mean that higher education is not
changing or not responding to societal demands or needs, it is just a reminder that
the sector is complex, and that the driving forces triggering change may sometimes
be contradictory or having side-effects which we are not aware of at this stage.
Certainly, the new media – of which rankings are an integrated part – are indeed a
new force to enter the field of higher education and with high impact potential on
policy-making. Still, the new media are not the only force, although they focus our
attention on the vertical dimension of institutional diversity (the best) and margin-
alise the horizontal one (the rest). But if we are entering the so-called knowledge
society, there is a need to cater to the characteristics and contributions of the whole
higher education system. Therefore, Teichler concludes, there is a need to improve
information in Europe about the impacts of the university experience in general,
and to develop alternative scenarios for the development of institutional patterns of
higher education systems which do not only embrace the ‘elite’ functions of the
system.

A RESEARCH AGENDA FOR RANKINGS

Although the full implications of rankings are yet unknown, rankings are probably
here to stay, and the pace in which new rankings are added to the already existing
ones suggests that we also will see many more rankings appear in the coming
years. As such, one could argue that rankings are still in their infancy, and that they
will probably become more sophisticated in the future. Research may still be
important in this process, and this book has identified a number of themes where
research may play an important role. First, and echoing much of the criticism related
to ranking methodology, there is a role for research to identify more relevant and
precise indicators of academic quality and the added value of higher education.
Although one can criticise rankings for the dominant use of input indicators, one
should perhaps also acknowledge that this emphasis most likely is due to a lack of
good and valid output measures. In this respect, there are already interesting initia-
tives taken in a number of countries and organisations to investigate how learning
outcomes and the added value of higher education can be measured and better
understood, and rankings would undoubtedly benefit from these efforts, if successful.
Second, and very much related to the first point, there is a need to link learning outcomes to the proper analytical level. Rankings have rightly been criticised for their tendency to aggregate data trying to rank institutions across departments and disciplines. Higher education research may be vital here because it can provide a bridge between the substance of disciplines and how this substance should be understood both academically and in a more policy-related context. In Europe, the so-called Tuning projects have generated new knowledge as to how we should characterise disciplines and study programmes across national and cultural borders, and future lines of research should take this initiatives one step further in trying to identify the common, acceptable or critical factors characterising a given discipline, academic area or study programme. It is most likely that such research would find much diversity and heterogeneity that could also pave the way for better information for those interested.

Third, and again following the former point, we need research that can help us to better understand how student and institutional behaviour are affected by rankings and similar information schemes. Much of the data that is provided today is based on what is thought to be important for students when making their choice where to study. However, as pointed out by several contributors in this book, there may be a number of cognitive, cultural and normative factors influencing student choice, and disclosing some of these may contribute to rankings that are of more relevance to students. Along the same lines, we need to know more about how institutions are affected by rankings, and the consequences of rankings on institutional actions. Here research should perhaps be particularly interested in the processes that lead up to strategic decisions studying on what grounds decisions are made, how information is acquired and interpreted, the weight given to rankings and such similar information. Such research could be of high relevance for both higher education institutions and the authorities regulating them. Not least, it could be of high relevance to understand what some authors in this book have labelled the new stratification of higher education.

Finally, rankings also address issues of accountability that we need to understand better. As mentioned earlier in this introduction, rankings may be seen as part of the accountability movement in society by which we all are affected. For higher education with its intangible products and services, accountability is a special challenge. This is probably why reputation is so important for many higher education institutions. Reputation can be used for accountability purposes, and can as such be an instrument of protection for those who enjoy high rating in academe and in society. But if reputation is a poor indicator of quality, the challenge is then to identify other variables and factors that can be used to illustrate the activities, profiles and contributions of the sector. Comparative research is needed to analyse accountability in different contexts, its links to cultural and cognitive characteristics of our societies, and to study how trust is created and maintained in higher education. Rankings are in many ways an indicator of distrust – the key question is whether this is possible to change, and what measures must be taken to accomplish this.

As we have pointed out, research is needed on a broad range of dimensions indicating that rankings actually offer a possibility to address some of the classical themes in higher education research: what is quality, how can we measure it, what
is the added value of higher education, and how higher education can best serve our societies. Certainly, those initiating the rankings should not be the only ones trying to answer these questions.

REFERENCES


PART I: METHODOLOGY
INTRODUCTION

University rankings or ‘league tables’, a novelty as recently as 15 years ago, are today a standard feature in most countries with large higher education systems. They were originally created over 20 years ago by US News and World Report in order to meet a perceived market need for more transparent, comparative data about educational institutions. Reviled by critics but popular with parents, copy-cat ranking systems began popping up all over the world, usually shortly after the introduction of – or a rapid rise in – tuition fees. Wherever rankings have appeared, they have been met with a mixture of public enthusiasm and institutional unease.

One of the main causes of institutional unease is the tendency of institutional ranking schemes to use weighted aggregates of indicators to arrive at a single, all-encompassing quality ‘score’, which in turn permits institutions to be ranked against one another. By selecting a particular set of indicators and assigning each a given weight, the authors of these rankings are imposing a specific definition of quality on the institutions being ranked. The fact that there may be other legitimate indicators or combinations of indicators is usually passed over in silence. To the reader, the author’s judgement is in effect final.

Intriguingly, however, there is little agreement among the authors of these indicators as to what indicates quality. The world’s main ranking systems bear little if any relationship to one another, using very different indicators and weightings to arrive at a measure of quality.

In an earlier paper (Usher and Savino 2006), one of the present authors discussed 17 sets of league tables from around the world. In this chapter, we update the Usher and Savino results by recording changes in methodology in a few of these rankings, as well as providing data on nine new systems of rankings. All told, twenty-two of these are ‘national’ league tables collected from fifteen countries (Australia, Canada, Chile, China, Hong Kong, Kazakhstan, Italy, the Netherlands, Peru, Poland, Spain, Taiwan, the Ukraine, the United Kingdom and the United States); while four are ‘international’ or ‘cross-national’ league tables. Specifically, we compare these league tables in terms of their methods of data collection and their selection and weighting of indicators. We also look at three other systems (the German CHE rankings, the SwissUp rankings and the Canadian University Navigator rankings produced by the Globe and Mail and the Educational Policy Institute) which do not conform to the standard league table ‘rules’.

© 2009 Sense Publishers. All rights reserved.
WHAT ARE UNIVERSITY RANKINGS AND LEAGUE TABLES?

University rankings are lists of certain groupings of institutions (usually, but not always, within a single national jurisdiction), comparatively ranked according to a common set of indicators in descending order. University rankings are usually presented in the format of a ‘league table’, much as sports teams in a single league are listed from best to worst according to the number of wins and losses they have achieved. Like performance indicators, they are made up of a series of individual indicators which are meant to present a statistical picture about a range of institutional qualities or activities. The primary differences are that while performance indicator systems simply present quantitative data on each indicator, league tables use them to emphasize differences between institutions, and in addition weight each indicator and then aggregate scores across a number of indicators to come up with a single ‘best’ institution.

Another notable aspect of league tables is that historically speaking, they have for the most part been produced by commercial publishing enterprises. In part, this reflects the fact that rankings share some characteristics with ‘consumer guides’ to various products. Although rankings are not guides to specific institutions, the publishers of individual institutional guides may incorporate rankings data as supplementary material, fleshing out descriptions for the purpose of providing more information to their readers. Rankings are – at least in theory – meant to be a way of helping citizens understand what they are getting for their public expenditures on education, and to help parents and students make informed decisions about where to spend their private dollars on education. However, as we shall see later, commercial rankers are increasingly being joined by governments who – in developing countries at least – seem to be eschewing performance indicator tables for rankings. Nigeria, Pakistan and Kazakhstan are among the countries where national governments have begun publishing rankings of their own national universities (Salmi 2007).

Institutional ranking systems can be conducted either on a national or international scale. National ranking systems are ones in which all or nearly all of a country’s universities are measured against one another. This was the original university ranking format – i.e., the type pioneered by the US News and World Report in 1981 and which has been widely copied in other countries. In most cases, all universities within a country are compared, although in some cases – notably in Canada (Maclean’s Magazine) and the United States (US News and World Report) – the country’s universities are divided up according to certain institutional characteristics and only compared to other institutions with similar characteristics, in effect creating a group of mini-league tables. It is rankings of these types that are included in this report.

Global institutional ranking systems are a new variation on the older idea of national rankings. There are at present three of these: the Academic Ranking of World Universities from Shanghai’s Jiao Tong University, first released in 2003, the World University Rankings from the Times Higher Education Supplement of Britain (henceforth THES), first released in November 2004, and the Ranking Iberoamericano rankings which compare universities throughout Latin America,
Spain and Portugal (a fourth which is not included in this survey is the Webometrics rankings, which measures institutional presence on the World Wide Web). A final international ranking considered here was one conducted by Asiaweek magazine prior to the publication’s discontinuation in 2000, which ranked the continent’s major universities.

Beyond institutional rankings, there are also sub-institutional rankings, which compare specific university units against similar ones at other institutions. These rankings are usually national in scope and deal with professional schools such as business, law and medicine. Graduate business schools are also the subject of a number of international rankings from such organizations as the Economist, the Financial Times, the Wall Street Journal and Business Week. These league tables were not covered in the 2006 Usher and Savino paper. However, one clear trend in European rankings is to issue institutional rankings entirely composed of rankings of departments (i.e. history, economics, biology), as is the case with the German CHE rankings, the Guardian rankings in the United Kingdom, the Italian La Repubblica and the two major Dutch rankings. We have therefore devoted some space in this review to what we called de-composed institutional rankings, which present a large number of sub-institutional rankings in a single format which does not necessarily result in an aggregate pan-institutional ranking.

There are also ranking schemes which focus on specific aspects of university activities. For instance, the Best American Research Universities ranks US institutions specifically on their research output, as, in a cruder manner, does the Centre for Science and Technology Studies in Bern, Switzerland, with its international ‘Champions League’ tables. Similarly, Yahoo Magazine has ranked universities on their ‘connectivity’, Webometrics on their internet presence and the Journal of Black Higher Education has graded them on their ability to integrate students from different backgrounds in its ethnic diversity rankings. These types of ranking systems are excluded from this survey because their purposes are much more specific and limited than the general ranking systems which we wish to focus on.

HOW RANKINGS AND LEAGUE TABLES WORK

League tables, by their very nature, are meant to boil down the work of entire institutions into single, comparable, numerical indicators. In most ranking systems, this comparison is a three-stage process: first, data is collected on indicators; second, the data for each indicator is scored; and, third, the scores from each indicator are weighted and aggregated.

All rankings systems operate by comparing institutions on a range of indicators. The number of indicators in a ranking system can vary significantly, from one in the simplest case (Banco Central de Chile, the Telegraph) to several dozen in the case of the most complicated (Excelencia). Specific areas of institutional activity or types of institutional output can therefore be compared across institutions, in much the same manner as is done with performance indicators.
With only a few exceptions (notably, Spain’s Excelencia rankings), league table systems then take the data on each indicator and turn it into a ‘score’. Usually, this is done by giving the institution with the highest score on a particular indicator a perfect mark of 100 and then awarding lower scores to other institutions based on how close they were to the score of the top institution.

Once scores have been derived for each indicator, they are weighted, with greater weight being accorded to indicators which are believed to be of greater importance. The weighted scores from all indicators are then tallied to give a unified final score for each institution. This weighting and aggregation is not always entirely transparent. There are examples like the United Kingdom’s Guardian and the Dutch Elsevier rankings where indicators do not possess specific weights; nonetheless, because institutions are still displayed from top to bottom, it is implicit that all indicators are being given an equal weight and the resulting scores are being aggregated in the normal way.

Clearly, the choice of indicators and the weight given to each indicator make an enormous amount of difference to the final output. Indeed, it is no exaggeration to say that when publishers advertise their product as a guide to ‘the best’ institutions, it is the publishers themselves who largely decide the best simply through their choice of indicators and weightings. In effect, the act of choosing a set of indicators and weightings imposes a one-size-fits-all definition of ‘quality’.

THE EVIDENTIARY BASIS OF LEAGUE TABLES – HOW DATA IS COLLECTED

A key issue in the preparation of league tables and rankings is the method by which data is collected. There are basically three sources of data on institutions:

- **Independent third parties.** Frequently, government agencies will collect and publish data on institutions in their jurisdiction, and this can be used as an objective standard by which to compare institutions. This data is very often financial in nature and is based on administrative data from grant-making bodies.

- **University sources.** The most complete and most detailed sources of data on universities are of course universities themselves, and they are thus potentially a very rich source of data.

- **Survey data.** Surveys of the opinions or experiences of various stakeholders can be used to obtain comparable data on different institutions regarding educational quality.

The use of each source of data has strengths and weaknesses. Independent third-party administrative data (usually from governments or grant-making bodies) is generally considered the ‘gold standard’ of comparative data since it is, at least theoretically, both accurate and impartial. The problem is that this data is not usually collected for the purpose of compiling league tables but rather as an administrative by-product of ordinary business. As a result, over-reliance on this source of data can lead to a situation where indicators are chosen simply on the basis that data is available rather than because they contribute to a sensible definition of quality. Certainly, there are few if any measures from these sources which
can provide much insight into the actual learning environments or even in many instances student outcomes from individual institutions.

It is for this specific reason, that many rankings systems – perhaps most notably *U.S. News and World Report* – have taken the trouble of asking institutions themselves to provide information about themselves for the rankings. In the United States, the result of so many news organizations doing this led to the creation of what is known as ‘the Common Data Set’, which has become a sort of alternative performance indicator datasource complementing the more narrow data available from the Department of Education and its IPEDS system. The benefit of this approach is that one can – in theory – answer a number of questions about institutional quality and learning environments that cannot otherwise be answered. For instance, one might be able to glean from government statistics what the teacher to pupil ratio at a school is, but one would not know things like average class sizes or the number of classes where professors (as opposed to graduate students) are teaching first year classes unless one asked institutions to provide the data themselves. The main drawback is that there is absolutely no guarantee that institutions will actually report the data to the ranker on a consistent basis, as all have a clear incentive to manipulate data in a manner which will benefit them. Indeed, at some institutions in the United States, there are staff positions within institutional research offices which require the incumbent to do nothing but provide institutional data to the *US News and World Report* in a favourable light. And, of course, as *Maclean’s* recently discovered to its cost, over-reliance on institutional data sources leaves one vulnerable to institutional boycotts.

Finally, there is survey data. Typically, surveys have been used to get the opinions of ‘experts’ – usually professors, administrators and employers – about the quality and reputation of various institutions. This data is scientific in the sense that it records observations accurately, but has been criticized for being misused or acting as a ‘lagging indicator’ of school quality. However, as Federkeil (2006) has shown, narrowly focussed reputation surveys can actually be very useful, as professor’s views about relative departmental quality is actually an extremely good proxy for research output. More recently, surveys of students have become an important source of data in rankings, notably in the German CHE, Canadian *University Navigator* and the two Dutch rankings. The increasing use of this type of data is creating a real shift in the kinds of topics which rankings are being used to assess.

Usher and Savino (2006) showed that surveys were the least frequently-used source of data among the various rankings examined; this was largely because the older surveys tended to only have one indicator (i.e. reputation) out of twenty or thirty populated by survey data. The CHE, Elsevier, *Keuzegids Hoger Onderwijs* and the *University Navigator* rankings all use survey data to populate half or more of all their indicators, so this has changed slightly (in addition, while THES uses surveys to populate only one out of its five indicators, but this indicator is worth 50% of the entire score).

Outside North America, third-party sources are by far the most heavily used sources of data: indeed, four of the 18 ranking schemes in this study use them exclusively. Of the remaining 14, third-party sources comprise a plurality of indicators in eight and university sources form a plurality in six. The predominance of
data from universities is most understandable in the cases of the *Asiaweek* and *THES* rankings, as their international scope significantly reduces the possibility of third-party sources providing data on a consistent trans-national basis. (The other two international rankings – *Shanghai Jiao Tong*, and *universia.com* – solved this problem by relying almost exclusively on research output measures such as scientific publications and citations.) In the cases of the *US News and World Report*, *Maclean’s*, the *Guardian* and *Rzeczpospolita*, the explanation seems to be that the editors’ definitions of ‘quality’ could not be measured using government administrative data. This may indicate a failure of government data collection in these countries, in the sense that information deemed important to quality measurement is not collected on a consistent and centralized basis; alternatively, it may indicate that the rankers’ views of what constitutes an indicator of quality is not shared by governments or the higher education community.

**WHAT LEAGUE TABLES MEASURE – A LOOK AT INDICATORS AND WEIGHTINGS**

It should come as no surprise to learn that different ranking systems use very different indicators in order to obtain a picture of ‘quality’. The number of individual indicators used in ranking systems worldwide runs well into the hundreds, making any kind of comparison grid too large to be useful.

In order to look at indicators and weightings in a manageable way, Savino and Usher categorized them into seven larger headings, based in part on an existing model of institutional quality written by Finnie and Usher (2005) (the alternative of using the four categories used by Dill and Soo (2004) was rejected because it was felt these categories were too broad). The Finnie-Usher work was originally a conceptual framework for quality measurement based on the following four elements:

- **Beginning characteristics**, which represent the characteristics, attributes and abilities of incoming students as they start their programs.
- **Learning inputs**, which come in two main types:
  - *resources*, both financial and material, available to students and faculty for educational ends; and
  - *staff*, both in terms of the numbers but also the way in which they are deployed to teach and the learning environment they create, as measured by the amount of contact time students have with their teachers, the kinds of exams they face, etc.
- **Learning outputs**, which represent the ‘skill sets’ or other attributes of graduates, which culminate from their educational experiences, such as critical thinking, analytic reasoning and technical knowledge. They also include records relating to retention and completion.
- **Final outcomes** represent the ultimate ends to which the educational system may contribute, including not only such traditional measures as employment rates and incomes but also any other outcome deemed to be important to individuals and society, such as job satisfaction, being a ‘good citizen’, etc.
These four elements were modified in the Usher and Savino (2006) study first, by making a clearer distinction between financial resources and staff and second by including two other sets of indicators, namely ‘research’ and ‘reputation’. This was an adequate encapsulation of the state of play as it was in 2006. Now in 2008, however, the field has moved on somewhat and some new categories of indicators seem to be worth discussing.

The first new set of possible indicators are those related to ‘internationalization’. A number of rankings, particularly in Europe (e.g. *la Repubblica* and *rzeczpospolita*) place a significant emphasis on institutional participation in Erasmus or on membership in other international research or institutional arrangements. The THES also puts significant emphasis on the number of international students and professors at an institution. For the most part, however, these indicators can remain categorized as ‘beginning characteristics’ or as one of the two forms of learning inputs, and as a result we have chosen not to create a new category yet, although as ranking proliferate some re-thinking may need to be done on this score.

The second new category of indicators has come about because of the increasing use of student surveys in the development of rankings. Many of the questions asked in these surveys effectively look at satisfaction about various aspects of institutional life. Where the questions ask directly about quality of teaching, or quality of resources (e.g. computers), we have categorized these variables as belonging to the relevant ‘learning inputs’ category. However, a number of the questions are not so easily categorizable, relating as they do to the perceived difficulty of classes, to the coherence of the material presented and the relevance of material presented to the world of work. In many ways, these questions are influenced by the kind of research that has over the years been summarized by Pascarella and Terenzini (2005), which focus on the importance of learning environments and student engagement as determinants of good outcomes. As a result, for this chapter we have added a new category of ‘Learning Environments’.

Rankings are, however, more than just a collection of indicators; instead, they are a weighted aggregation of indicators. It is therefore important to see how they are put together and how each ranking system implicitly defines educational quality through the distribution of its weighting. Although the apparent differences between ranking systems are substantial, it turns out that there are some real and intriguing similarities among particular subsets of league tables.

Table 1, below, shows the differences in the indicators and weightings used by different league table systems. Each row summarizes the distribution of indicator weightings among the seven categories of indicators described in the previous section and adds up to 100 per cent. It is obvious from even the most cursory glance at this table that no two ranking systems are alike and indeed that some have virtually no areas of overlap with one another.
Table 1. League Table Weightings (rankings added since Usher and Savino (2006) in bold; rankings which have changed methodology since 2006 are indicated in italics.)

<table>
<thead>
<tr>
<th>Institution and Region</th>
<th>Beginning characteristics</th>
<th>Learning inputs – staff</th>
<th>Learning inputs – resources</th>
<th>Learning Environment</th>
<th>Learning outputs</th>
<th>Final outcomes</th>
<th>Research</th>
<th>Reputation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Melbourne Institute – International Standing of Australian Universities – Australia (year?)</td>
<td>11</td>
<td>3.5</td>
<td>11</td>
<td>0</td>
<td>12.6</td>
<td>4.8</td>
<td>40</td>
<td>17.1</td>
</tr>
<tr>
<td>Guangdong Institute of Management Science China (2007)</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>57.1</td>
<td>0</td>
<td>42.9</td>
<td>0</td>
</tr>
<tr>
<td>Netbig – China (2007)</td>
<td>12</td>
<td>21.8</td>
<td>6</td>
<td>0</td>
<td>0</td>
<td>45.2</td>
<td>15</td>
<td></td>
</tr>
<tr>
<td>Shanghai Jiao Tong University World Rankings China/World (2007)</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>10</td>
<td>0</td>
<td>90</td>
<td>0</td>
</tr>
<tr>
<td>Wuhan University Centre for Science Evaluation – China (2007)</td>
<td>10.6</td>
<td>8.5</td>
<td>16.6</td>
<td>0</td>
<td>3.4</td>
<td>0.6</td>
<td>48.6</td>
<td>11.7</td>
</tr>
<tr>
<td>Education18.com – Hong Kong (2007)</td>
<td>20</td>
<td>10</td>
<td>10</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>30</td>
<td>30</td>
</tr>
<tr>
<td>Asiaweek – Asia’s Best Universities (2000)</td>
<td>25</td>
<td>28.3</td>
<td>10</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>16.7</td>
<td>20</td>
</tr>
<tr>
<td>La Repubblica – Italy (2007)</td>
<td>16.7</td>
<td>31.4</td>
<td>21.9</td>
<td>0</td>
<td>10.0</td>
<td>0</td>
<td>20</td>
<td>0</td>
</tr>
<tr>
<td>Perspektywy/Rzeczpospolita Uniwersytet Europe – Poland (2007)</td>
<td>17.5</td>
<td>26</td>
<td>31.5</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>25</td>
</tr>
<tr>
<td>Excelencia, – (2001)</td>
<td>0</td>
<td>25</td>
<td>25</td>
<td>0</td>
<td>25</td>
<td>0</td>
<td>25</td>
<td>0</td>
</tr>
<tr>
<td>Daily Telegraph UK (2003)</td>
<td>0</td>
<td>100</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Financial Times UK (2003)</td>
<td>9</td>
<td>19</td>
<td>15</td>
<td>0</td>
<td>10</td>
<td>27</td>
<td>20</td>
<td>0</td>
</tr>
<tr>
<td>Guardian University Guide UK (2007)</td>
<td>17</td>
<td>32</td>
<td>17</td>
<td>0</td>
<td>17</td>
<td>17</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>The Times Good University Guide – UK (2007)</td>
<td>3.3</td>
<td>53.3</td>
<td>6.7</td>
<td>0</td>
<td>3.3</td>
<td>3.3</td>
<td>30</td>
<td>0</td>
</tr>
<tr>
<td>THES World University Rankings UK/World (2007)</td>
<td>5</td>
<td>25</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>20</td>
<td>50</td>
</tr>
</tbody>
</table>
The 26 rankings displayed here could scarcely be more different. Some are heavily weighted towards a single type of indicator (research in the case of the Shanghai and Iberoamericano rankings, teaching in the case of the Telegraph rankings, and graduate salaries in the case of the Banco Central de Chile); others spread their rankings across a number of indicator areas. Despite the vastly different choices of indicators and weightings evident throughout the world, certain patterns do appear when the studies are grouped together geographically. For instance, studies from China – which has four different ranking projects – place much more weight on research indicators than any other studies in the world. In the most extreme case – that of Shanghai Jiao Tong University’s Academic Ranking of World Universities – research performance is worth 90 per cent of the total ranking. This is followed by Wuhan, where research measures are worth 48.2 per cent of the final ranking, Netbig (45.2%), and Guangdong (42.1%). Much of this weighting comes from counting papers and citations in bibliometric studies – studies which have a heavy bias towards the hard sciences. In contrast, comparatively little weight is put on either resource inputs or on final outcomes. Whether this is because data on these issues is scarce or because Chinese experts genuinely consider indicators of these types to be unimportant is an open question.
Other regional patterns are also evident. Rankings of UK universities, for instance, completely eschew the use of reputation surveys as a means of determining quality (although THES places a 50% weighting on reputation issues). British league tables also put a much higher emphasis than league tables elsewhere on measures of staff and staff quality – on average, they put over 40 per cent of their weighting in this area, as opposed to an average of just five per cent in the rest of the world’s league tables combined. The two Dutch league tables mirror each other almost exactly – with a major emphasis on Learning Environment factors and the rest on Learning Inputs.

Seven sets of rankings – La repubblica, Rzeczypospolita, Excellencia, the Times, Maclean’s, Keuzegids Hoger Onderwijs and the Ukrainian university rankings – put over 50 per cent of the weighting on the two types of learning input factors. Intriguingly, neither of the two rankings from the United States – the country where disparity in institutional resources is presumably the greatest – deems it fit to put much weight on these two indicator areas.

Reputation seems to be a kind of all-or-nothing affair. Most ranking systems do not use reputation as an indicator; however, where it is used, it tends to be weighted quite heavily. Learning Input – Staff is the most commonly used indicator type, with 21 of 26 rankings systems using them. But the most consistently highly weighted area in rankings systems is research performance and research output, which is used in 18 of the 26 rankings and, on average, accounts for over a third of the weight in those rankings.

Table 1 graphically demonstrates the central premise of this chapter: different ranking systems have very different definitions of quality. The notion of ‘quality’ in higher education is clearly a very malleable one – some observers wish to look at outputs, while others focus on inputs. Among both inputs and outputs, there is very little agreement as to what kinds of inputs and outputs are important. Not only is no single indicator used across all ranking schemes, no single category of indicators is common either: remarkably, none of the seven basic categories of indicators are common to all university ranking systems.

One of the only previous comparative examinations of league tables (Dill and Soo 2004) concluded, on the basis of an examination of four sets of league tables in four countries, that international definitions of quality were converging. Our findings, based on a larger sample, contradict their result. We acknowledge that part of the reason for the contradiction lies in the fact that we have divided indicators into seven categories instead of four and hence were always likely to find more variation. Methodological differences notwithstanding – and we believe our methodology to be the more refined of the two – the results still conflict. We believe that had Dill and Soo looked at Asian or international ranking schemes, they too would have seen these differences and revised their conclusions.

CONSISTENCY OF OUTCOMES ACROSS LEAGUE TABLES

One might reasonably conclude from the foregoing analysis that measured institutional quality is not immutable and that an institution’s ranking is largely a function of what the ranking body chooses to measure. A possible example in support of
this proposition is Queen’s University in Kingston, Canada. In its domestic rankings (Maclean’s), it fares very well because it attracts good students and is reasonably well-endowed and well-funded. In international rankings, it fares poorly, even compared to other Canadian universities, because its small size puts it at a disadvantage in terms of non-normalized research output measures.

Due to the plethora of ranking systems that have appeared in recent years, one can now test this proposition directly. In most countries, there are at least three separate rankings ‘observations’ made by different national and international ranking systems (those of THES and Shanghai Jiao Tong, plus one or more domestic rankings). In those instance where one can use multiple ranking schemes to look at the relative scores of institutions in a single country, we find that certain institutions invariably rise to the top: Oxford and Cambridge in the United Kingdom; Harvard, Yale, Princeton, MIT and Stanford in the United States; Peking and Tsinghua in China; and the University of Toronto in Canada. Despite the very different weighting and aggregation schemes used by the domestic and international league tables, these institutions manage to consistently monopolize the top spots. Further down the ordinal ladder, the different rankings systems start to show greater variation (i.e. there is rarely any agreement between systems as to which university lies in tenth position) but regardless of the ranking scheme employed, ‘top universities’ almost always seem to come out as top universities.

This poses a serious problem for interpretation. If rankings were absolutely inconsistent across all league tables, it would be easy to dismiss the whole idea of ranking as an intellectually worthless exercise designed simply to sell newspapers or magazines. If rankings were absolutely consistent across all league tables, then we might conclude that there are probably one or two ‘super’ indicators which are driving the overall rankings, with the remainder of the indicators essentially being ‘chaff’ with which to distract readers and to create false differentiations. But neither of these scenarios is true – in fact, what appears to happen is that different ranking schemes provide consistent results for some institutions and inconsistent ones for others.

The simplest explanation for this is that institutional league tables don’t measure what their authors think they are measuring. League tables’ authors believe that each indicator is a reasonable proxy for quality and that, suitably aggregated and weighted, these indicators constitute a plausible, holistic ‘definition’ of quality. In fact, most indicators are probably epiphenomena of some underlying feature that is not being measured. That is to say, there is actually some ‘dark matter’ exerting a gravitational pull on all ranking schemes such that certain institutions or types of institutions (the Harvards, Oxfords and Tsinghuas of the world) rise to the top regardless of the specific indicators and weightings used. A search for this ‘dark matter’ certainly seems deserving of future research. Our guess, however, is that ‘age of institution’, ‘faculty size’ and ‘per-student expenditure’ are probably excellent candidates to be this ‘dark matter’.
RANKINGS WITHOUT LEAGUE TABLES: THE CHE APPROACH

For most of this chapter we have been describing league tables – that is, ranking systems that provide a single integrated score that allows an ordinal ranking of entire institutions. However, this is not the only possible approach to university rankings. There is, for instance, no intrinsic reason why indicators must focus solely on institutions; approaches which look at institutions at lower administrative levels (such as departments or faculties) are also possible. As we noted earlier, this is the approach taken by the Guardian, la Repubblica and the two Dutch rankings. These systems provide comprehensive departmental-level rankings across entire universities (that is to say, they provide separate rankings for each discipline).

A different approach altogether has been taken by the Centre for Higher Education Development (CHE) in Germany which issues annual rankings jointly with a media partner (currently Die Zeit, formerly Stern). The CHE has also been responsible for producing the SwissUp rankings in Switzerland, and its approach has been copied in North America by the University Navigator.

There are five significant features to the CHE approach to rankings. The first is that rankings are done at a sub-institutional (i.e. field of study) level (although the University Navigator differs from the others in this respect as it is done at an institutional level). The second is that no weights are attached to any of the indicators. The third is that absolute scores are not given in any of the indicators: institutions are grouped roughly into the thirds, and displayed as being in the top, middle, or bottom tiers. The fourth is that the indicators are not aggregated to create an overall ‘score’. As a result, no institution is declared ‘best’; the indicators of results are merely displayed. Fifth, the results are online and users of the rankings may in effect create their own rankings by selecting the indicators which are of interest to them and then receiving personalized results based on their choices.

It should be noted that not all of these five features are unique to CHE-style rankings. The first two features, for instance, are shared by The guardian, La Repubblica, Elsevier and the Keuzegids Hoger Onderwijs. Putting rankings online and making them susceptible to manipulation is not unique to these three rankings, either: in Canada, Maclean’s has begun to allow readers to re-jig the rankings on their own on their online portal, though the main magazine still adheres to its old format of weighting and aggregating indicators.

While none of these individual factors is necessarily a radical departure from established practice in rankings, taken together, they do constitute a fairly radical departure in the concept of rankings. While they are certainly ‘rankings’ in the broad sense, they are not ‘league tables’ because they do not declare one institution as better than another on the basis of aggregate scores. CHE avoids this precisely because it believes that it is at best meaningless (and at worst actively misleading) to combine widely disparate indicators into a single overall hierarchy.

This stance presents certain difficulties in presenting data in a printed format. Instead of a simple ordinal rank, all indicators must be shown for all institutions, which means that they are somewhat unwieldy and difficult to read. On the other hand, this stance has an enormous advantage when translated to the World Wide Web, where it is possible for users themselves to in effect create their own weightings...
and rankings by selecting a restricted number of indicators and asking the web site’s database to provide comparative institutional information on that basis. In so doing, the CHE approach effectively cedes the power of defining ‘quality’ – which, as we have seen, is one of the key roles arrogated by the authors of ranking schemes – to consumers of the ranking system (i.e. prospective university students and their parents or sponsors).

Though it is not a logical corollary of the CHE approach, it is instructive to note that the three main interactive rankings are also fairly distinctive when it comes to data sources and the types of indicators used. In all three cases, massive numbers of students are surveyed about their study experiences and levels of satisfaction at their university. In Germany, the CHE conducts regular surveys of approximately 130,000 students and 16,000 faculty, covering nearly 250 higher education institutes. In Canada, about 25,000 students are surveyed each year. In all three cases, between half and two-thirds of indicators are based on surveys of students and – as is the case in the Elsevier and Keuzegids Hoger Onderwijs rankings – the result is a reasonably strong emphasis on Learning Experience variables. These are shown below in table 2.

### Table 2. Distribution of Indicators in Interactive ‘CHE-Style’ Rankings

<table>
<thead>
<tr>
<th></th>
<th>Beginning characteristics</th>
<th>Learning inputs – staff</th>
<th>Learning inputs – resources</th>
<th>Learning Environment</th>
<th>Learning outputs</th>
<th>Final outcomes</th>
<th>Research</th>
<th>Reputation</th>
</tr>
</thead>
<tbody>
<tr>
<td>CHE – Germany (2007) – 21 indicators**</td>
<td>1</td>
<td>1</td>
<td>10</td>
<td>5</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>SwissUp – Switzerland (2007) – 36 indicators**</td>
<td>3</td>
<td>4</td>
<td>9</td>
<td>9</td>
<td>1</td>
<td>2</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>University Navigator – Canada (2007) – 27 indicators</td>
<td>2</td>
<td>5</td>
<td>12</td>
<td>4</td>
<td>0</td>
<td>0</td>
<td>4</td>
<td>0</td>
</tr>
</tbody>
</table>

** in both the CHE and SwissUp rankings, the exact indicators used differ somewhat from subject to subject. For SwissUp we have excluded four indicators which appeared to be used only for medical faculties; in the case of CHE, where there is quite significant variation from one subject to another, we took the social sciences categories as a base. In both cases, we have excluded an indicator relating to rent levels in towns surrounding the university as there was no way to fit these into the overall model.

It is tempting to see the CHE-style rankings as being a distinct alternative to League Tables and in many ways they are. But on the other hand, we also see that a number of rankings – especially the two Dutch rankings – share a great deal in common with the CHE-style rankings while still superficially resembling US News-style League Tables. Clearly, there is not a single immutable model of university rankings; many approaches are possible and indeed many approaches are
currently being undertaken. In all cases, the basic tools – indicators populated by quantitative measurements – are the same. The biggest difference between rankers is philosophical; traditional rankings with weightings aggregation by definition are trying to find an institution which can be called ‘the best’; CHE-style rankers are using the power of the internet to help consumers find the institutions that is ‘best for them’.

CONCLUSIONS

Based on this survey of league tables, we can conclude a number of things, notably:

– There are vast differences between university league tables in terms of what they measure, how they measure it and how they implicitly define ‘quality’.

– Some of these differences appear to be geographic or cultural in nature. There is notable clustering of certain types of indicators and certain types of data sources. Whether this reflects genuine differences in opinion about the definition of what constitutes ‘quality’ in universities or cross-national differences in the collection and availability of data is unclear, although we lean towards the former explanation. The lack of common indicators across countries explains why the large international league tables (*Shanghai Jiao Tong* and *THES*) are so reliant on measures of publication outputs and on reputational surveys (respectively), as they are the only indicators that do not rely on governments or institutions to first collect and process the data.

– An increasing number of rankings are using student survey data to populate their rankings and as a result there is an increasing bias towards looking at the student experience. While this may seem ideal in some quarters, it is worth stating that very little research has been done in order to ensure that answers to these responses are in fact apples-to-apples. *Perception* of quality is a very different metric than quantitative measurements such as ratios of professors to students or size of research budgets. Most significantly, if students at school A have significantly higher expectations of service and quality than students at school B, then it is very difficult to interpret results on any questions related to satisfaction. While data about the student experience of the learning environment are surely very useful and important to potential applicants, the comparability of these data across institutions for the purpose of objective ranking is still unclear.

– Despite major inconsistencies in the methodologies used to rank universities, there is a surprising level of agreement between ranking systems as to which universities in a given country are ‘the best’. To the extent that different methodologies give differing opinions about the quality of an institution, the variance between observations grows as one moves down the ordinal rankings.

– Although the definition of ‘quality’ is contested, league tables by definition impose a ‘one-size-fits-all’ approach to the matter; this is precisely why they are so controversial. As the CHE approach shows, however, league tables are not the only way to approach rankings. Indeed, the spread of the World Wide Web provides collectors of institutional data with an opportunity to democratize rankings.
and put the power of ranking in the hands of the consumer by following an ‘any-size-fits-all’ approach.

As Merisotis (2002) has noted, university rankings are here to stay. As imperfect as they are, they satisfy a public demand for transparency and information that institutions and governments have not been able to meet on their own. Moreover, as higher education becomes more costly for individuals and families, the demand for comparative information on universities will increase. As a means of delivering that information, however, league tables are only in their infancy, and all of them can clearly benefit from greater analysis of the assumptions implicit in their own schemes. This is particularly the case with respect to international league tables, which have a restricted range of possible indicators due to the lack of available cross-national comparative data. To the extent that international ranking schemes are taking on a quality assurance role in the growing international student market, this suggests that the global higher education community needs to begin to look at how best to collect and report data on institutions so as to permit thoughtful and responsible inter-institutional comparisons.

NOTES

1 An organization with which both of the present authors are associated.
2 The term stems from UK-based chart listings that were often compared with Premier League professional soccer or football standings in England during the 1990s and can now be found in an extremely wide variety of contexts in Britain today.

REFERENCES


NEW RANKINGS CITED IN THIS REPORT


GERO FEDERKEIL

REPUTATION INDICATORS IN RANKINGS
OF HIGHER EDUCATION INSTITUTIONS

INTRODUCTION
In the context of growing competitiveness among higher education institutions worldwide, national as well as international rankings of universities have assumed added significance both for higher educational bodies as well as for the general public. Rankings have a strong impact on institutions: in most countries, students’ decisions to opt for an institution are influenced, to varying degrees, by rankings; political decisions about higher education might also be based on rankings. In a few cases, accreditation or funding is also related, either directly or indirectly, to rankings. One of the most important effects of rankings is their impact on the reputation of institutions, both at a national and an international level. As early as 1993, Theus pointed out, that “as resource-acquiring institutions, colleges and universities depend upon positive public attitudes to attract the students, faculty and financial resources” (1993, p. 278). In particular, the two global rankings published during the last years contributed widely to the international ‘race for reputation’ and the reputational hierarchies of universities. Reputational indicators play a significant role in rankings. According to Giddens’ notion of ‘duality of structure’ (1984), there is what might be called a ‘duality of reputation’: Reputation comprises of both the outcome as well as the medium of rankings. On the one hand, reputational hierarchies are measured by rankings; on the other, rankings influence reputational hierarchies by that very measurement. In some rankings, reputational indicators substantially contribute to the final score; yet, they are among the most controversial of issues when it comes to rankings (cf. Dill and Soo 2005, p. 503).

Although the use of reputational indicators is criticised within higher education institutions, research shows that certain target groups (e.g. prospective students) are interested in knowing those universities that have a commendable reputation. A German study showed that 52 per cent of prospective students select a university because of its high reputation and it is one of the most common factors in the decision-making process (HIS 2005, p. 198). Unpublished data of the CHE ranking also suggest that reputation is an important factor for students when selecting a university. Furthermore, ranking data show that reputation is more important for engineering students than for students of the humanities. Therefore, universities compete for students by referring to high reputation on their web sites. Last but not least we have to take into account that data on reputation can easily be collected by surveys among groups of stakeholders.

© 2009 Sense Publishers. All rights reserved.
REPUTATION INDICATORS IN RANKINGS

While existing rankings vary enormously with regard to their purposes, target groups, methodology and indicators of reputation are used in most rankings in some form or the other. Reputation indicators had already been used in some ranking procedures as early as the 1960s, as, e.g., in a ranking of U.S. doctoral programmes published in 1966 by the American Council on Education (Cartter 1966). And they are also part of most current rankings.

In one of the two existing global rankings, the Times Higher Education Supplement World Rankings (THES), two indicators of reputation are used that together make up for 50 per cent of the overall score. The first one refers to reputation among scholars, and the second to reputation among employers. In the most prominent national U.S. ranking, the U.S. News & World Report Ranking, information about the reputation of colleges and graduate schools has a weight of 25 per cent in the total ranking score. Other rankings that use indicators of reputation are. rankings made by the Polish magazine Perspektivy (50% of the total score), Asia Week (25%), Macleans (16%) and the CHE (but no weight as there is no overall score). The most notable exception is the Jiaotong Ranking that completely relies on factual indicators for their research (as bibliometric indicators and Nobel prices).

Table 1. Reputation indicators in rankings

<table>
<thead>
<tr>
<th>Ranking</th>
<th>Share in total score</th>
</tr>
</thead>
<tbody>
<tr>
<td>THES World Rankings</td>
<td>50%</td>
</tr>
<tr>
<td>U.S. News</td>
<td>25%</td>
</tr>
<tr>
<td>Perspektivy (Poland)</td>
<td>50%</td>
</tr>
<tr>
<td>Asia Week</td>
<td>25%</td>
</tr>
<tr>
<td>Macleans</td>
<td>25%</td>
</tr>
<tr>
<td>CHE</td>
<td>Not available*</td>
</tr>
</tbody>
</table>

* The CHE ranking does not calculate an overall score.

The concepts and the operationalisation of reputation differ between the rankings: Some rankings refer to the reputation within the academic community (e.g. THES, CHE), while others refer to employers (the THES second indicator, some business school rankings). The most common approach is the one where a survey is conducted among a particular group of respondents who are asked to name a limited number of institutions they think to be the best – either institutions as a whole or specific fields/disciplines in the institution. Only very few rankings ask respondents to order all universities in a field and/or region according to their reputation (e.g. a reputational study of U.S. undergraduate education published in 1981 by Solmon and Astin). With a strict methodology on the basis of the most common ‘top group’ approach, only a top group could be calculated, but not a complete league table of the reputation of all institutions in a given context: If respondents
are asked to name the top five institutions, a university that is regarded to be number six by every single respondent would have a zero count – contrary to a university that most respondents would rank at the bottom but that is mentioned among the top five just by a very small number of respondents. On the other hand, it is quite questionable that respondents could give an assessment of every single university in a regional or field sample that could be the basis for a complete reputational hierarchy.

**THE CONCEPT OF REPUTATION**

In his analysis of the social system of science, Luhmann stresses the importance of reputation for the operation of the system and describes reputation as the “second selective code” (1992, p. 352) in the system of science/higher education besides the basic distinction of false–true as the guiding difference in the system of science. “Reputation on the positive side refers to the first communication of new knowledge and, on the negative side, to the lack of that achievement” (ibid.). In complex systems like science there is a problem of ‘overcharging observers’ with information (Luhmann 1992, p. 245). In this situation the reference to symbols – like reputation – allows ‘a shortened orientation’ that is reducing the complexity of information. The plausibility of reputation, however, depends from the fact that “the hand that is allocating reputation, remains ‘invisible’” (Luhmann 1992, p. 246); a quasi-official allocation of reputation would give it a political dimension. Although rankings have no official status (with very few exceptions) they make reputational hierarchies public and hence the hand becomes visible – this might explain why reputation indicators are particularly disputed among the academic community itself.

By using reputation in rankings, the information must be transformed into an indicator. According to a common definition, indicators “provide a measurement for assessing the quantitative or qualitative performance of a system” (Cuenin 1986, p. 7). But reputation indicators are not indicators of performance. In general usage the notion of reputation refers to “a characteristic or attribute ascribed to one person (firm, industry, etc.) by another” (Wilson 1985, p. 27). Reputation is also described as a general social attitude of quality (Theus 1993, p. 282), that reflects a ‘kaleidoscope of attitudes’ and factors ranging from research performance to, at least in the United States, athletic powers. This indicates that different groups of stakeholders (e.g. students, parents, academics, employers, politicians) have different conceptions when speaking about the reputation of higher education institutions. This means reputation refers to the social ascription of some – positively evaluated – characteristics by particular groups of people. In the context of higher education, reputation refers to the social ascription of high competence, good performance or excellence. In a conceptual manner the relation between reputation and actual performance is contingent.

This ascription of reputation can be made differently by various social groups and stakeholders; hence the reputation of an institution can differ by those groups. Different groups have different criteria in ascribing reputation to an institution. Hence there is no reputation as such, rather reputation refers to a particular social group.
In particular, reputation of higher education institutions might differ
- between different social groups and different stakeholders; e.g. the reputational
  hierarchy of employers may be different from that of higher education policy
  makers or the academic community,
- between different fields of science; a university with a high reputation in physics
  may have a sociology department with a low reputation and
- according to regional and national aspects which is of particular relevance to
  international rankings.

At the same time reputation of institutions among particular stakeholders and in
particular fields might be affected and biased by something like an overall reputation
of institutions. This refers to the ‘halo’-effects (Thorndike 1920), a cognitive bias
whereby the perception of a particular institution is influenced by the perception of
the former traits or traits in a different setting in a sequence of interpretations. With
regard to reputation in rankings, this means that a department may receive higher
evaluations when they are found within institutions that have, on the whole, a high
reputation (Brooks 2005, p. 7). An example for this can be found in a ranking made
by a German business magazine in which employers rated business studies
programmes; they rated Heidelberg University, which generally has a high reputa-
tion, among the top six, but that University does not have a business studies
programme (see also Fairweather 1988; Clarke 2002). In chapter 3 those aspects
are analysed empirically by using data from various rankings.

Using the concept of Bourdieu, reputation can be understood as a form of social
capital that can be used in a competitive field. Bourdieu defines social capital as
“the aggregate of the actual or potential resources which are linked to possession of
a durable network of more or less institutionalized relationships of mutual
acquaintance and recognition” (Bourdieu 1983, p. 190). This means that this capital
is inherited from the past and must be continuously reproduced and is dependent on
the inclusion in social networks. Both aspects play a role in determining the reputa-
tion of higher education institutions. The ascription of reputation refers to the
perception of achievements in the past but has to be continuously reproduced so as
not to turn into a relict of ‘glory from the past’. And, reputation as social capital
can be transformed into economic capital, e.g. in the context of the allocation of
research funds.

THE SOCIAL CONTEXTUALITY OF REPUTATION

As a social ascription, reputation of higher education institutions and the reputa-
tional hierarchy of institutions might be differing between social groups or groups
of stakeholders.

Reputation among Different Social Groups

As reputation is the social ascription of excellence/high performance it might differ
between different social groups or stakeholders who can have different criteria for
ascribing reputation or have a different perception of ‘good performance’. To illustrate
this aspect we use the example of two German rankings using reputation indicators but within different groups. The CHE ranking gives information on the reputation among professors of an academic field (cf. Müller-Böling and Federkeil 2007), a ranking published by the magazine *Karriere* (Career) asked about the reputation among a sample of the biggest 1,000 German employers. Both rankings refer to the reputation in business education programmes with regard to education and training. If we compare the results of both studies we see that reputational hierarchy of both groups is very similar with regard to German public universities. But whereas employers ascribe a high reputation to some small private business schools, those business schools have almost no reputation at all among German professors of business education. This example illustrates that different groups can have different opinions on reputation of universities due to different rationales: Whereas employers probably have a priority on issues of ‘employability’, this does not have much priority for professors who rather think of research or academic quality when taking reputation into consideration.

**Reputation by Academic Disciplines**

Most rankings (such as THES World Rankings, U.S. News) report about reputation of universities as a whole. Clearly there is some understanding of the reputation of universities as a whole – everyone thinks of Harvard, Oxford or Cambridge when thinking of the world’s top universities. But reputation can be very different for different academic fields and most users of rankings are interested in one particular subject – students who want to find a university for a programme as well as researchers who want to inform about the universities in their field.

To give an example with data from the CHE ranking: The field-specific reputation indicators in the CHE ranking are based on a survey among all professors of a field. Professors are asked to list up to five universities they believe to be the leading universities in their field (except their own) – separately for research and for studying a degree programme (cf. Berghoff et al. 2006). The indicator gives the percentage of votes of a university. The reputation, e.g., of Humboldt University in Berlin, is quite different in different academic fields included in the ranking. Evidence shows big differences between reputation scores of different departments. The range with regard to reputation in research varies from three per cent to more than 30 per cent. Whereas computer sciences and chemistry at Humboldt University have almost no reputation at all among German professors of those fields, the Medical and the History Departments belong to the faculties with the highest reputation in Germany. Those numbers indicate that the reputation does not only differ between broad fields (like sciences and humanities) but even within broad fields between single disciplines (e.g. English and Romance languages or Mathematics vs. computer sciences).
Figure 1. Reputation by disciplines – Example Humboldt University Berlin

Again, this example shows that reputation is highly dependent on social groups ascribing reputation to an institution – in this case, members of different academic communities. To put it in statistical terms: reputation is highly dependent on the sample asked about reputation: The more the medical or history professors in the survey the higher the reputation of Humboldt University would be, the more computer scientists the lower. Calculating the overall reputation of Humboldt University across disciplines would then – even in a sample controlled for disciplinary structure – blur the differences between disciplines and would inevitably show a university with some middle reputation. To sum up, reputational indicators are highly sensitive to the field structure of the sample.

**International Reputation**

The last aspect refers to international rankings. Reputation has to do with knowledge about an institution – in the extreme, with knowledge about the very existence of a university. Normally high reputation focuses on a small number of institutions. An example from the CHE ranking (sociology in 2005) shows that more than 80 per cent of all votes by German sociology professors concerning the reputation of their departments in research focuses on only eleven universities (out of 39; see illustration 2).
REPUTATION INDICATORS IN RANKINGS OF HIGHER EDUCATION INSTITUTIONS

Figure 2. Concentration of high reputation (CHE ranking sociology)

Figure 3. Reputation of German and Swiss universities (CHE ranking, physics)
The relationship between reputation and knowledge about institutions is particularly evident in international contexts. The most prominent example of a ranking using reputation as an indicator for international comparison is the THES World Ranking. The THES ranking on the one hand asked a sample of academics throughout the world about the leading universities; they should state for which field and which region they feel competent. On the other hand it refers to an international survey conducted among recruiters. Unfortunately, in both cases there is no information about the exact return rates and the sample structure of both the surveys – neither in terms of disciplines nor in terms of countries/regions of the world. So there is no possibility to make a more detailed analysis by fields and regions with the THES ranking data.

Data from the CHE ranking, which started to internationalize in 2004 by including the universities of Austria and Switzerland, can illustrate the problems related to reputation indicators in international rankings. Since then the sample for the reputation survey includes professors from Germany, Austria and Switzerland; they were asked about the leading universities (up to five) in the three countries. Hence a comparison of reputational hierarchies among professors in these three countries is possible. Our example compares the reputation of German and Swiss universities in physics among German and Swiss professors (cf. figure 1). We see that German and Swiss professors pretty much agree about the reputation of German universities. But whereas there is a clear and graded hierarchy of Swiss universities among Swiss professors, the Swiss physics departments have – with the only exception of the ETH Zürich – almost no reputation at all among German professors.

So there is clear evidence that an international – or even global – reputational ranking of universities is heavily dependent on the sample included in the survey in terms of regions/nationalities. Referring to Bourdieu’s idea that reputation as social capital is linked to networks being continuously reproduced, we see that such networks notwithstanding processes of globalisation still have a spatial/regional dimension.

REPUTATION AND LEAGUE TABLE APPROACHES

In most rankings, reputation indicators – as well as a total score for the institution/department – are transformed into a league table with clear-cut ordinal positions. Evidence from the CHE ranking and other rankings suggest that the distribution of reputation, which is usually expressed as a percentage of votings, is extremely uneven: Very few institutions make up for a very big percentage of votings. Accordingly the numerical differences of reputation scores among the lower ranked institutions are rather small.
The problems of transforming such a skew distribution into league tables can be illustrated by changes in reputational ranking over time in the THES ranking. If we conceive reputation as an expression of social networks we can expect reputation scores/hierarchies to remain rather stable over time – in particular with regard to the one-year intervals of many rankings. By comparing the original reputation scores of the THES ranking in 2004 and 2005 we find a high correlation ($r = .93$; cf. fig. 4). Similar results could be shown with CHE ranking data.

Transforming those original reputation scores into league table positions, the fallacies of the league table approach become evident: In a sample of 200 universities almost two-third moved by more than ten rank positions (and 41% even more than 20 positions) within a single year when original scores are strongly correlated. This example indicates that league table approaches tend to exaggerate the differences between universities and changes over time and produce artefacts.
REPUTATION AND (RESEARCH) PERFORMANCE

Since their beginnings, reputation indicators have been criticised for their lack of objectivity and validity. Conceptually, reputation is not linked to performance; the relationship between the notion of reputation and performance can be seen as contingent. Cave et al. (1997, p. 173) report that there can be large and dramatic differences between reputation indicators and ratings based on objective measurements of productivity. Due to the fact that reputation refers to the perception of – past – performance by people ascribing reputation, there might be universities that are overestimated as well as underestimated with regard to their actual performance. A survey of reputation studies in the United States concluded “that reputational scores demonstrate a high overall consistency with other program measures” (Brooks 2005, p. 7).

CHE ranking data show that there are quite considerable correlations between performance indicators and reputation. But the size of the correlation varies between disciplines: for example, the correlations in Sociology are weaker than
in mechanical engineering. Furthermore, within disciplines correlations are different for different indicators. In some disciplines reputation has a strong impact on the ability of the universities to get external research grants, in others to the number of PhDs and in some to the number of publications.

<table>
<thead>
<tr>
<th>Discipline</th>
<th>Research funds per year</th>
<th>Publications per year</th>
<th>Citations per publication</th>
<th>Number of PhDs per year</th>
</tr>
</thead>
<tbody>
<tr>
<td>Biology</td>
<td>.65**</td>
<td>.73**</td>
<td>.62**</td>
<td>.65**</td>
</tr>
<tr>
<td>English Studies</td>
<td>.64**</td>
<td>.66**</td>
<td>n.a.</td>
<td>.57**</td>
</tr>
<tr>
<td>Business Studies</td>
<td>.41**</td>
<td>.53**</td>
<td>n.a.</td>
<td>.65**</td>
</tr>
<tr>
<td>Chemistry</td>
<td>.65**</td>
<td>.69**</td>
<td>.34**</td>
<td>.59**</td>
</tr>
<tr>
<td>Economics</td>
<td>.49**</td>
<td>.44**/.82**</td>
<td>n.a.</td>
<td>.32*</td>
</tr>
<tr>
<td>Electrical Engineering</td>
<td>.75**</td>
<td>n.a.</td>
<td>n.a.</td>
<td>.86**</td>
</tr>
<tr>
<td>Education Studies</td>
<td>.78**</td>
<td>.60**</td>
<td>n.a.</td>
<td>.49**</td>
</tr>
<tr>
<td>History</td>
<td>.65**</td>
<td>.69**</td>
<td>n.a.</td>
<td>.70**</td>
</tr>
<tr>
<td>Mechanical Engineering</td>
<td>.80**</td>
<td>n.a.</td>
<td>n.a.</td>
<td>.87**</td>
</tr>
<tr>
<td>Mathematics</td>
<td>.64**</td>
<td>.59**</td>
<td>n.a.</td>
<td>.60**</td>
</tr>
<tr>
<td>Medicine</td>
<td>.65**</td>
<td>.56**</td>
<td>.42**</td>
<td>.59**</td>
</tr>
<tr>
<td>Physics</td>
<td>.61**</td>
<td>.71**</td>
<td>.54**</td>
<td>.81**</td>
</tr>
<tr>
<td>Psychology</td>
<td>.61**</td>
<td>.33*</td>
<td>.16</td>
<td>.46**</td>
</tr>
<tr>
<td>Sociology</td>
<td>.56**</td>
<td>.61**</td>
<td>n.a.</td>
<td>.50**</td>
</tr>
</tbody>
</table>

n.a. – not available

Figure 6. Correlation reputation in research and performance indicators (CHE 2006)

Again, those differences are meant to distinguish between disciplines and not to compare universities as a whole. Correspondingly, the correlations between research performance – measured by the number of publications – and reputation are much smaller ($r = .35$) in the THES World Rankings of whole universities (cf. Figure 7).
Figure 7. Reputation and Performance in THES World Rankings

But even given the relatively high general correlations between reputation and research performance indicators in the CHE ranking there are universities that do not fit into this picture. Looking at the correlation between publications and reputation in mathematics ($r = .59$) we see universities with high performance and low reputation and vice versa, i.e. there are some underestimated as well as overestimated universities if we look at real performance (cf. Figure 8). Overestimation may be related to good performance in the past whereas underestimation may be a problem for new institutions without long traditions and hence without the possibility to be included in relevant networks, which is a major factor in gaining reputation.

This complex dynamics of performance and social ascription is closely linked to a temporal dimension. New insights are valued as a precondition for further innovation and are hence rewarded with reputation (Luhmann 1992, p. 250) – “No system of reputation could survive if reputation would be allocated arbitrary or undeserved in many instances” (ibid., p. 251). But Luhmann points out that this implies an effect of exaggerating following the premise ‘once good, always good’. But he rightly says that this does not occur without any factual basis, as the past reputation helps to get resources and funds, better positions and better possibilities to publish (p. 251). In this sense reputation “is working self-reinforcing: Once there, it has a better chance to grow”. Only in rare instances reputation is ‘de-temporalised’ – either with the ‘unrecognised genius’ (the underestimated in fig. 8) who would have deserved reputation earlier than he does or what Luhmann calls “the classique”
who still has high reputation even if his “work has lost its importance for research” (p. 251). In this context, van Raan (2005, p. 95) points to the fact “that established reputation is not necessarily the same as ‘past glory’. Often we see that institutions with an established reputation are strong in maintaining their position. They simply have the best possibilities to attract the best people, and this mechanism provides these renowned institutions with a cumulative advantage to further reinforce their research performance”.

Figure 8. Reputation and Performance – CHE ranking

Analysis of US News & World Report ranking data indicate that correlates much stronger with high research expenditures than with good graduation-rate performance (Graham and Thompson 2001) and confirm CHE ranking data showing that reputation is stronger related to research than to teaching performance.

CONCLUSIONS

Although controversial, reputation indicators are used widely in rankings; in some they contribute substantially to the overall ranking score of universities. In rankings, but also outside the world of rankings, reputation is an important information about
universities. According to Luhmann reputation can be seen as the “second selective code” operating within the system of science that is helping to reduce the complexity of the system. Reputation may become a social fact with regard to recruiting of students and scientists and to funding decisions, too. Reputation is a form of social capital within the system of higher education that can be transformed into economic capital, too.

Analytically, reputation is conceived as the social ascription of high performance/good quality to institutions by individuals. In analytic terms it must be distinguished from performance. There are some examples of halo-effects that universities might be considered among the best in a particular discipline that they do not have at all because the institution as a whole has a high reputation. Empirical study shows that there are considerable correlations between academic performance and reputation within academic communities but linear correlation coefficients blur a number of outliers – universities that are overestimated or underestimated in reputation with regard to their actual performance.

As a social ascription reputation is highly dependent on the social context and on social groups. Empirical evidence shows that reputation differs between different groups of stakeholders (e.g. employers and academics); i.e. it is different for different disciplines and fields of universities and, last but not least, reputation is dependent on social networks and hence – even in global academic communities – differing with regard to regional and national aspects. There is no reputation as such, there is only reputation among a particular sample of people.

Those findings suggest that reputation indicators can be a valid information in rankings – but only as an information on reputational hierarchies within defined social groups, not as a measure of performance or quality. But validity of reputation measures depends on clear information about the sample (reputation among whom? reputation where?). A deconstruction of reputation myths of institutions is only possible if a ranking gives the relevant information on reputation and at the same time reputation can be contrasted by performance indicators. Calculating an overall ranking score including both reputation and performance indicators inevitably blurs the differentiation between performance and reputation. As rankings contribute to the reputation of universities including the very information on reputation into an overall score (with high weights in some instances, like, e.g., the THES World Rankings) or even exclusively relying on reputation indicators means to construct a circular argument. Using reputation indicators in such a way is to reify the social construction of reputation.

NOTES

1 First, the “Academic ranking of world universities” published by the Shanghai Jiaotong University (first in 2003); see: http://ed.sjtu.edu.cn/ranking.htm; and second the “World University Rankings” by the Times Higher Education Supplement, first published in 2004, see: http://www.thes.co.uk/worldrankings/.

2 Currently there is some unverified information that the return rate in the most recent ranking was only one percent!
REFERENCES


