The Changing Academy – The Changing Academic Profession in International Comparative Perspective 3

Jung Cheol Shin
Robert K. Toutkoushian
Ulrich Teichler *Editors*

University Rankings

Theoretical Basis, Methodology and Impacts on Global Higher Education

Springer

University Rankings

The Changing Academy – The Changing Academic Profession in International Comparative Perspective 3

Scope of the series

As the landscape of higher education has in recent years undergone significant changes, so correspondingly have the backgrounds, specializations, expectations and work roles of academic staff. The Academy is expected to be more professional in teaching, more productive in research and more entrepreneurial in everything. Some of the changes involved have raised questions about the attractiveness of an academic career for today's graduates. At the same time, knowledge has come to be identified as the most vital resource of contemporary societies.

The Changing Academy series examines the nature and extent of the changes experienced by the academic profession in recent years. It explores both the reasons for and the consequences of these changes. It considers the implications of the changes for the attractiveness of the academic profession as a career and for the ability of the academic community to contribute to the further development of knowledge societies and the attainment of national goals. It makes comparisons on these matters between different national higher education systems, institutional types, disciplines and generations of academics, drawing initially on available data-sets and qualitative research studies with special emphasis on the recent twenty nation survey of the Changing Academic Profession. Among the themes featured will be:

1. Relevance of the Academy's Work
2. Internationalization of the Academy
3. Current Governance and Management, particularly as perceived by the Academy
4. Commitment of the Academy

The audience includes researchers in higher education, sociology of education and political science studies; university managers and administrators; national and institutional policy-makers; officials and staff at governments and organizations, e.g. the World Bank.

For further volumes:
http://www.springer.com/series/8668

Jung Cheol Shin • Robert K. Toutkoushian
Ulrich Teichler
Editors

University Rankings

Theoretical Basis, Methodology and Impacts on Global Higher Education

Springer

Editors
Jung Cheol Shin
Seoul National University
Department of Education
Gwanak-ro 599
151-742 Seoul, Gwanak-gu
Korea, Republic of (South Korea)
jcs6205@snu.ac.kr

Robert K. Toutkoushian
University of Georgia
Institute of Higher Education
Meigs Hall
30602 Athens Georgia
USA
rtoutkou@uga.edu

Ulrich Teichler
Universität Kassel
Internationales Zentrum für
Hochschulforschung Kassel
Mönchebergstr. 17
34109 Kassel
Germany
teichler@incher.uni-kassel.de

ISBN 978-94-007-1115-0 e-ISBN 978-94-007-1116-7
DOI 10.1007/978-94-007-1116-7
Springer Dordrecht Heidelberg London New York

Library of Congress Control Number: 2011928551

Preface

The book is designed to provide the reader with a comprehensive understanding on university ranking schemes – its theoretical basis, methodological issues, and impacts on society. Although rankings have been used in policy and academic discussions, there are rare books in the market which provide the comprehensive theoretical and methodological issues on rankings itself. Recently, policymakers and the media often criticize their universities based on their ranking status. Further, many of institutional leaders set rankings as their benchmark in their vision or master plan for the university. Nevertheless, academics are quite skeptical of rankings because they believe that they mislead higher education institutions as well as have enormous methodological limitations.

The book is not about how to enhance the ranking status of an institution, nor how to devise better ranking systems. Instead, we focused on how to widen the understanding of university rankings for different audiences – academics, rankers, and general people mostly parents and students. Academics, rankers, and their audiences have different views on university rankings and use them for different purposes. A group of academics have been developing precise technical methods of measuring institutional performance, but are less interested generally in its social impacts. Rankers do not pay much attention to the impacts that ranking games bring on higher education society. Instead, they are interested in how to update their ranking schemes better than their competing rankers so that they can sell the rankings through their publications. Audiences of rankings only pay attention to the ranking status of the university they are interested in, but not to the details behind the rankings. This book is designed to help bridge this gap between ranking specialists (academics), rankers, and ranking audiences.

The book has three major parts. Part I reviews the theoretical and practical basis of university rankings, Part II focuses on the methodology used to derive rankings, and Part III discusses the social impacts of university rankings.

In Part I, Jung Cheol Shin begins Chap. 2 with a discussion of organization effectiveness, which is the theoretical base of ranking survey, neglecting academic or policy discussions in many cases. Although a university's ranking is a dimension of measuring institutional effectiveness, rankers and its audiences often disregard the fundamental issue that there are other aspects of effectiveness that should also be considered. Since the 1980s, the organizational effectiveness of higher education

institutions has been evaluated by governments or their agencies to directly or indirectly link institutional performance with budget allocation in some countries (e.g., the USA, Australia, the UK, etc.). In Chap. 3, Grant Harman discusses university rankings as well as public accountability and quality assurance as an aspect of organizational effectiveness. Finally, in Chap. 4, Ulrich Teichler introduces and discusses ranking literature focusing on theoretical and practical agendas of rankings and its impacts on higher education.

In Part II, we discuss the methodologies used to derive university rankings. In Chap. 5, Bernard Longden shows how rankers easily shift rankings by changing weights or by inclusion or exclusion of indicators. In Chap. 6, Karen Webber provides a comprehensive overview of how to measure institutional research, teaching, and service performance. In the next chapter, Robert K. Toutkoushian, Karen Webber, and Keith Trigwell discuss the details of measuring research and teaching performance. In Chap. 8, Lutz Bornmann leads further discussions on measuring research performance focusing on indicators, peer review, and reputation. This chapter provides contemporary theoretical and practical issues of measuring research performance, such as how to count citations in the real world. The last topic we cover in Chap. 9 of Part II is about teaching quality.

In Part III, the respective contributors discuss the social impacts of ranking survey. Although rankings are a dimension of organizational effectiveness, the social impacts of rankings are enormous. Part III begins with social influences of ranking survey at institution level. In Chap. 10, Christopher Morphew and Christopher Swanson discuss how higher education institutions respond to rankings. They have focused on how universities try to enhance their ranking status. In the Chap. 11, William Locke discusses how universities institutionalize ranking schemes into their internal systems and cultures. In Chap. 12, Akira Arimoto goes further on the ranking impacts on higher education institutions and discusses how the rankings impact faculty life on campus.

The contributors of the book are well-known researchers in higher education worldwide and have many years of teaching and research experiences at higher education institutions. We are confident that the issues we highlight and the theory and practice issues we discuss will contribute to academic society of ranking studies and also to the development of higher education. The work on this book can be traced back to papers that were prepared for and presented at the International Conference on Education Research (ICER-11), which was hosted by Education Research Institute, Seoul National University in October 2009. We are grateful to Education Research Institute for the financial supports and organizing the conference. We thank Jung-Eun Lee and Hyun-Ju Park graduate students at Seoul National University for their contribution to the editorial works.

Seoul National University, South Korea — Jung Cheol Shin
University of Georgia, USA — Robert K. Toutkoushian
University of Kassel, Germany — Ulrich Teichler

Acknowledgment

Partial fulfillment of this work was supported by the National Research Foundation of Korea Grant funded by the Korean Government (NRF-2010-330-B00232).

Contents

Contributors

Jung Cheol Shin. Jung Cheol Shin is associate professor at Seoul National University. His research interests are higher education policy, knowledge production and social development, academic profession, quality assurance, and higher education finance. His researches have been published in international journals such as *Higher Education*, *Review of Higher Education*, *Studies in Higher Education*, *Education Policy Analysis Archive*, and *Asia Pacific Education Review*. He is actively involving in international journal editorial boards. He is the Executive Editor of *Asia Pacific Education Review* and the Review Editor of *Higher Education*. He has been invited as guest speakers by leading research universities in many countries including the USA, Japan, China, Australia, Mexico, etc. Currently, he is leading a worldwide research on knowledge production and social development with Prof. Loet Leydesdorff of University of Amsterdam and Prof. Han Woo Park of Yong Nam University funded by National Research Foundation of Korea.

Robert K. Toutkoushian. Robert K. Toutkoushian is professor in the Institute of Higher Education at the University of Georgia. His areas of specialization include the economics of higher education, higher education finance, labor market issues in higher education, and the use of quantitative research methods in higher education. He obtained a Ph.D. in economics from Indiana University in 1991, and has worked at the University of Minnesota, University System of New Hampshire, and Indiana University prior to joining the Institute of Higher Education in July 2009. Dr. Toutkoushian served as Editor-in-Chief of the series *New Directions for Institutional Research*, and currently he is Editor-in-Chief for the journal *Research in Higher Education*. He has published approximately 30 studies in peer-reviewed academic journals and 15 chapters in edited books, and has edited volumes on the economics of higher education and faculty compensation.

Ulrich Teichler. Ulrich Teichler is professor at the University of Kassel since 1978, former director for altogether 16 years of INCHER-Kassel, former dean of faculty, and vice president of the University of Kassel (for 2 years). His research areas are higher education and the world of work, international comparison of higher education systems, and international cooperation and mobility in higher education, and the academic profession. Former or current co-editor of various academic journals, e.g., *Higher Education*, *European Journal of Education*. He has

done consultancy activities for UNESCO, OECD, World Bank, Council of Europe, and European Commission. He is a former chairman of the Consortium of Higher Education Researchers (CHER) (1988/1989 and 1992–1998), former president (1998–2002) and distinguished member of European Association for Institutional Research (EAIR). He published more than 1,000 academic publications. His recent publications include *Higher Education Systems* (2007), *Universities as Centres of Research and Knowledge Creation: An Endangered Species?*, *Higher Education and the World of Work* (2009), and *The Professional Value of ERASMUS Mobility* (co-author, 2009).

Akira Arimoto. Akira Arimoto is professor and director of Higher Education at the Research Institute for Higher Education, Hijiyama University, and Professor Emeritus of Hiroshima University, Japan. He is Associate Member of the Japan Council of Science, and President of the National Association of RIHE. He is a member of editorial board of the book series *Higher Education Dynamic*. He was Professor and Director of the Research Institute for Higher Education at Hiroshima University. He served Chair of UNESCO's Global Scientific Committee for the Asian and Pacific Region, President of the Japanese Association of Higher Education Research (JAHER), and President of the Japan Society of Educational Sociology (JSES). He was a visiting fellow to Yale University, Max Planck Institute, and Lancaster University, as the first Nitobe Fellow at the International House of Japan. Recent publications include books as *National Innovation and the Academic Research Enterprise* (ed. D. Dill and F. Van Vught), *Introductory Educational Sociology*, *Academic Profession and FD*, and *The Changing Academic Profession in Japan.*

Bernard Longden. Professor Longden has been engaged with the higher education sector for over 35 years – initially as a lecturer in biological science and more recent years, as a researcher on higher education policy. Over the past 10 years he has developed his research capacity and encompassed issues around the student experience, e.g., student non-completion identifying barriers preventing students from a successful outcome to their program of study. Similarly, he has published work on performance measures, rankings and league tables, and challenging the assumptions on the nature and quality of data used in the production of commercial publications. More recently, he has analyzed qualitative and quantitative data from over 25 universities to explore the nature of the first year student experience. Bernard Longden was Chair of the European Association for Institutional Research, and is currently the Editor-in-Chief of *Tertiary Education & Management*.

Christopher C. Morphew. Christopher Morphew is professor of Higher Education and chair of the Department of Educational Policy and Leadership Studies at the University of Iowa. In 2009, Professor Morphew's co-edited volume (with P. Eckel), *The Privatization of the Public Research University,* was published by Johns Hopkins University Press. His scholarly articles have appeared in many competitive journals such as the *Journal of Higher Education*, *Research in Higher Education, Review of Higher Education*, and *Higher Education Policy*. Professor Morphew is

a member of the Executive Board for the Association for the Study of Higher Education and was previously an associate of the National Center for Public Policy in Higher Education, a Salzburg Seminar fellow, and a Western Interstate Commission for Higher Education/Ford Foundation Policy Fellow. He has built an international reputation as manifested in invited international presentations in China, Brazil, Portugal, Norway, the United Kingdom, South Korea, and the Netherlands.

Christopher Swanson. Christopher Swanson is a doctoral student in the Department of Educational Policy and Leadership Studies at the University of Iowa. His research interests center around social stratification in education opportunities and outcomes, credentialism and its implications on institutional policy, and the gender gap reversal in education attainment.

Grant Harman. Grant Harman is emeritus professor and director of the Centre for Higher Education Management and Policy at the University of New England (UNE). He is former chair of the UNE Academic Senate and Pro Vice-Chancellor (Research). His main research interests are in higher education policy, quality assurance, knowledge transfer and commercialization, the academic profession, and international education. He is Editor-in-Chief of the Springer journal, *Higher Education*, and has worked as a consultant for various Government Ministries in Australia and internationally for the World Bank, the Asian Development Bank, UNESCO, and the South African Vice-Chancellors' Association. He is a co-editor of the recent volume *Higher Education Reform in Vietnam: Challenges and Priorities*, published by Springer.

Karen L. Webber. Karen Webber is associate professor and graduate program coordinator in the Institute of Higher Education at the University of Georgia. Prior to joining the IHE faculty on a full-time basis, she served as Director of Institutional Research and Interim Associate Provost for Institutional Effectiveness at the University of Georgia. Her primary research interests are related to faculty productivity and the assessment of academic, cognitive, and psychosocial growth of college students, with additional interests in gender studies and higher education data management. She teaches doctoral level courses on Institutional Research, Research Methods in Higher Education, and The NCES/NSF National Datasets. She is actively involved in the Association for Institutional Research (AIR), The Society for College and University Planning (SCUP), and the Association for the Study of Higher Education (ASHE).

Keith Trigwell. Keith Trigwell is professor of Higher Education in the Institute for Teaching and Learning (ITL) at the University of Sydney. He was previously Director of the ITL and of the Oxford Centre for Excellence in Preparing for Academic Practice, a Fellow of Kellogg College and Reader in Higher Education at the University of Oxford. His research interests include investigating qualitative differences in university teaching and students' learning experiences, teaching–research relations and the scholarship of teaching, including development of the Approaches to Teaching Inventory. This work has yielded more than 100 journal articles, conference papers, and books, including *Understanding Learning and*

Teaching: The Experience in Higher Education, which summarizes 10 years of research with Professor Michael Prosser. He is a former president of the International Society for the Scholarship of Teaching and Learning, and since 2005, has been an associate editor for the international journal *Higher Education*.

Lutz Bornmann. Since the late 1990s, Dr. habil. Lutz Bornmann has been working on issues in the promotion of young academics and scientists in the sciences and on quality assurance in higher education. His current academic interests include research evaluation, peer review and bibliometric indicators, especially the *h index*. He is a member of the editorial board of the Journal of Informetrics (Elsevier) and of the advisory editorial board of EMBO Reports (Nature Publishing group). Since 2004, he has published more than 50 papers in journals covered by Thomson Reuters with a total of more than 500 citations. His *h index* amounts to 12 (the publication list and citation metrics are available on http://www.researcherid.com/rid/A-3926-2008).

William Locke. William Locke directed the UK part of the international study of the *Changing Academic Profession* from 2006 to 2010, while Assistant Director of the Centre for Higher Education Research and Information (CHERI) at the Open University. He has edited two books and published several academic journal articles, reports, and chapters, including: the higher education research–policy nexus; intermediary bodies in higher education; markets in UK higher education; mergers and organizational cultures in higher education institutions; and the strategic management of the research–teaching nexus. He has spoken at international conferences in Australia, China, Japan, North America, and throughout Europe. He has written a number of policy reports, including student engagement; graduates' retrospective views of their courses; conceptions of excellence in teaching and student learning; and the career decision-making of part-time students. Since writing this chapter, William has become Head of Learning and Teaching at the Higher Education Funding Council for England (HEFCE).

Chapter 1
The Past, Present, and Future of University Rankings

Jung Cheol Shin and Robert K. Toutkoushian

In this chapter, we provide the groundwork for the entire book. Although we do not discuss each topic in detail, the intention is to convey introductory information for readers about the topics to be covered by the various contributors. In the section dealing with university rankings in higher education contexts, we briefly review the development of ranking surveys and introduce the concept of organizational effectiveness, discuss the concepts of quality and quantity in higher education, and the mechanisms that are used to measure organizational effectiveness. In the methodology section, we introduce the reader to measures of institutional performance and related issues. In the section dealing with the impacts of ranking on society, we focus on the impacts of ranking surveys on higher education systems, individual institutions, students, and the side effects of ranking surveys. We close this chapter by discussing the future of ranking surveys.

1.1 University Ranking Within Higher Education Contexts

There is a long history of colleges and universities competing with each other for students, teachers, donors, and social support. For a long time, the competition has been evaluated by implicit reputation without any data to back up perceptions. However, with the heightened competition between universities since the 1990s and the dramatic growth of the international higher education market, surveys have emerged in many countries as a means of evaluating and ranking universities. Recently, the competition has been accelerated in many countries as governments

J.C. Shin (✉)
Department of Education, Seoul National University, Seoul, South Korea
e-mail: jcs6205@snu.ac.kr

R.K. Toutkoushian
Institute of Higher Education, University of Georgia, Athens, GA, USA
e-mail: rtoutkou@uga.edu

J.C. Shin et al. (eds.), *University Rankings*, The Changing Academy – The Changing Academic Profession in International Comparative Perspective 3,
DOI 10.1007/978-94-007-1116-7_1,

develop initiatives to build world-class universities that can compete more effectively with other leading institutions across the globe. Although there are concerns with using rankings as a tool for measuring the quality of a university, many institutional leaders and policymakers still often rely on rankings to inform their policymaking.

Ranking universities is a challenging task because each institution has its own particular mission, focus and can offer different academic programs. Institutions can also differ in size and have varying amounts of resources at their disposal. In addition, each country has its own history and higher education system which can impact the structure of their colleges and universities and how they compare to others. It is therefore very difficult to rank entire universities, especially across national borders, according to the single criterion of ranking indicators. An alternative approach that has been used by some is to, instead, rank academic programs, with early pioneers including Hughes (1925, 1934), Cartter (1966), and Roose and Andersen (1970).

The practice of ranking university graduate programs started in 1925 when Professor Donald Hughes (1925) ranked graduate programs on the basis of peer reputations in the United States. Since then, many initiatives to rank graduate programs and institutions have been tried by academic researchers and by research institutes (Ehrenberg and Hurst 1996; Cartter 1966; Conrad and Blackburn 1985; Drew and Karpf 1981). For example, beginning in 1982, the National Research Council (NRC) has periodically collected information on research-doctorate programs in 274 institutions and 41 disciplines in the United States (Webster and Skinner 1996).

Compiling rankings of institutions presents challenges to academic researchers. Following the graduate program rankings done in 1925 by Donald Hughes, it would be 60 more years before any ranking of institutions was undertaken. In 1983, the US News and World Report (US News) started its college rankings and gradually expanded its focus from undergraduate to graduate education and from ranking institutions to ranking graduate programs.

Following the example of the US News ranking, a growing number of commercial media and research institutions have begun to release ranking worldwide and/or nationally. Some of the most well-known international ranking schemes include the Academic Ranking of World Universities by Shanghai Jiao Tong University, the Times Higher Education Supplement-QS World University ranking (Times QS ranking), the Leiden University ranking, and the Taiwan Higher Education and the Accreditation Council ranking. In their comprehensive overview of rankings, Usher and Medow (2009) reported that at the time of their study, there were a minimum of 26 rankings worldwide. In addition, new ranking systems have been emerging in many countries. In our web search of ranking surveys for this book, for example, we found that as of 2009, there were at least 33 ranking systems of higher education around the world.

The rapid growth of university rankings might be explained by the increasing interest in egalitarianism in higher education. The concept of egalitarianism in higher education competes with the elitism ideal and argues that higher education

should focus on providing services to the general population, as well as the elites. This ideal emerged in the late nineteenth century in the United States with the rapid massification of higher education. Since then, different types of higher education institutions such as community colleges in the United States, polytechs in Europe, and other types of two-year institutions have emerged in many countries as a means of increasing egalitarianism.

Similarly, with the rapid growth of higher education markets, policymakers and employers began to raise the issue of quality in the 1980s. Elite universities soon began to compete with each other to attract better qualified students and to attract financial resources from donors. The general public also began to be interested in the activities and accomplishments of universities and how they compared to each other. This societal interest led to the emergence of ranking systems for higher education. At the same time, policymakers began to establish quality assurance schemes to comply with the questions of quality issue of mass higher education. Another accountability effort came from public sector management with the liberal governments in the United States and the UK in the 1980s developing public accountability systems. This required higher education institutions to report their performance to policymakers according to predetermined performance indicators. Currently, these three different mechanisms—*rankings* developed mainly by the media, *quality assurance measures* created by quality assurance agencies, and *accountability measures* imposed by governments—co-exist in higher education.

Why have university rankings become so popular with the policymakers and institutional leaders since the mid-1990s? Some have argued that the growing interest in ranking is closely related to the massification, marketization, and globalization of higher education (Shin and Harman 2009; Dill 2009). Rankings can help consumers see the value of their investment in higher education and hold institutions accountable for results. Rankings can provide students with comparisons of institutions in different countries. In addition, socio-political contexts, neo-liberalism, also add up the popularity of ranking among policymakers as well as parents and students.

1.2 Theoretical Basis of University Rankings

Occasionally, the raters of universities and the consumers of rankings do not pay much attention to the fact that rankings were initially conceived as a tool for measuring the effectiveness of higher education institutions. It is often assumed that highly ranked institutions are more productive, have higher quality teaching and research, and contribute more to society than lower-ranked institutions. However, the three main dimensions of institutions – teaching, research, and service – can differ or even conflict each other, and thus institutions that are performing well in one area may perform poorly along another dimension. For example, a small institution may be very efficient in educating a given number of students with limited

resources, but not very efficient in the production of research. On the other hand, we might find a large institution that is very efficient in knowledge production but not in teaching undergraduate students.

It has been suggested that a production framework can be used to describe how universities transform inputs into outputs (Astin 1993; Cave et al. 1991; Borden and Bottrill 1994). Drawing on this theory, an organization's effectiveness can therefore be measured by a combination of inputs, throughputs, and/or outputs. For example, organizational effectiveness can be measured by goal attainments (outputs), efficiency of goal attainment (e.g., inputs vs. outputs), or other combinations of these elements.

University ranking surveys tend to focus on all three aspects at the same time. For example, the US News ranking relies on input measures (financial resources, alumni giving, entrée students academic achievement, etc.), and outputs and outcomes (retention, graduation rates, program ratings of peers). This means that a highly ranked university is one that invests in more resources, and thus in turn produces better results. However, some rankings consider only metrics in one of these areas and disregard the others. This omission is contrary to the key theoretical issue for the notion of "productivity," which is how to maximize the production of outputs given the resource inputs.

Some initiatives have been conducted by academic researchers to consider relative efficiency by considering both inputs and the outputs produced. For example, Zheng and Stewart (2002) ranked universities in the United States by the relative efficiency which considers input as well as outputs of higher education institutions. As the pioneering studies in this line of inquiry show, program and institutional rankings based on different perspectives on organizational effectiveness can produce results that are quite different from those in the more popular rankings.

Another theoretical issue that should be addressed is how to properly measure the concept of "quality" in higher education. Because we rarely have good measures of the true outputs from higher education (such as the quality of learning and research), it is very difficult to find proxy measures for this construct. For the reason, a critical issue for researchers is how to use the available quantitative indicators to best represent quality in higher education. Institutions typically measure teaching quality based on student course evaluations, and measure faculty research productivity by their publication counts, citation counts, and/or level of sponsored research funding.

Do these measures really capture the quality aspect of teaching and research? Although these measures might represent quality to some degree, we believe that they represent only limited dimensions of quality of teaching and research. For example, a scholar who publishes many articles might be considered to be very productive; however, the scholar's research may in fact contribute less to knowledge than his/her colleagues who published fewer, but more significant, articles. Similarly, a faculty member with high teaching evaluations may in fact be less effective than other faculty at conveying knowledge and helping students to learn the subject matter and ultimately succeed in college.

Most of the popular university rankings, especially global rankings, rely heavily on quantifiable measures of institutional performance. Other rankings such as those produced by the *Times QS*, however, also use measures based on reputation surveys of academic colleagues and business employers. The *National Research Council* (US) and *US News and World Report* also use reputation measures to evaluate graduate programs. Although reputation measures have limitations in reflecting the quality of teaching and research (van Raan 2007), an argument can be made that a program's reputation is the accumulated performance over a long time, and is therefore more stable and reliable than most of quantified measures and pulls together the various aspects of productivity and quality (Johnes 1988).

As noted earlier, there are three types of mechanisms that can be used to measure organizational effectiveness: rankings, accountability measures, and quality assurance. Rankings are led by for-profit media or individual research institutes and tend to attract significant attention from policymakers. Unlike rankings, quality assurance and accountability measures are led by government agencies. These three mechanisms share common goals, however, namely the improvement of quality in higher education.

Accountability initiatives are most often developed by governments to ensure higher education institutions are providing the right services to the public. Most accountability systems have indicators which then allow the government to judge whether the institutions have attained predetermined goals. Sometimes, accountability systems utilize financial benefits or sanctions to help and/or entice institutions to respond to the initiatives (Shin 2010). For example, performance-based budgeting programs carry financial benefits while performance reporting does not. The aim of quality assurance initiatives is to ensure higher education institutions provide a minimum level of quality in education. Many countries established accountability systems in the 1990s in an effort to help enhance the quality of higher education. Quality assurance systems also rely on indicators, but tend to emphasize qualitative data obtained though external peer reviews, and may be conducted by a separate quality assurance agency. Quality assurance emphasizes quality improvement rather than how an institution compares to others based on designated metrics. For this reason, quality assurance might be described as being more university friendly, whereas rankings are student friendly and accountability initiatives are policymaker friendly.

1.3 Methodological Issues in Evaluating Quality

The question we pose in this section is how to practically measure institutional quality. The concept of quality of a university should apply to all three primary functions of an institution: teaching, research, and service. To implement this, metrics for each must be developed and decisions made about how the metrics should be weighted and combined.

1.3.1 Dimensions of University Rankings

University rankings are typically based on some combination of institutional performance (research, teaching, services, etc.), institutional characteristics (institutional mission, size, regional locations, etc.) and other factors. Of these, institutional performance is the stated consideration of many ranking surveys, especially rankings that compare institutions across nations. In the discussion below, we cover some of the methodological issues that must be addressed in the various ranking methods that are in use today.

1.3.1.1 Measuring Teaching Quality

Occasionally, teaching quality is measured by statistics generated from student class evaluations, although there are controversies on how to properly measure teacher quality. Some researchers have argued that teaching quality should be measured by learning outcomes, student attitudes, and the behaviors changed through their college education (Brew and Boud 1995; Elton 2001; Simons and Elen 2007). However, in practice it is quite difficult to measure learning outcomes or students' improved competency due to the instruction that they have received. As a result, researchers and higher education institutions may have little choice but to rely on course evaluation surveys as a proxy measure for teaching quality. Although some researchers such as Feldman (1989), and Marsh (1987, 2007) believe that course evaluations are useful because they are highly correlated with student learning outcomes, there is still controversy within academe over whether they should be used to measure teaching quality.

Many institutional ranking schemes do not incorporate course evaluations into their rankings because each institution uses its own course evaluation process, and the data from these evaluations are rarely made available to others, which means that the ranking agency cannot directly compare different institutions. To overcome this, rankers may rely on metrics that they believe are correlated with teaching quality, such as the faculty-student ratio and the expenditures per pupil for instruction. However, these input measures do not guarantee quality of teaching though the inputs provide better teaching environments.

Some rankers also have developed a market mechanism to evaluate teaching quality. For example, the Times QS rankings use the results from an employer satisfaction survey to serve as a proxy for teaching quality. This method is an attempt to measure teaching quality from the customer's perspectives. Another noteworthy approach to capture the quality of teaching can be found in student college experience surveys such as the National Survey on Student Engagement (NSSE) in the United States, College Student Experience Questionnaire in Australia, and Teaching Quality Assessments in the UK. College experience surveys enable rankers to use consistent data to compare higher education institutions. Both employer surveys and college experience surveys have also been used to demonstrate accountability

as well as produce rankings of institutions. Nonetheless, rankings generally rely more heavily on measures of research quality than on teaching quality.

1.3.1.2 Measuring Research Quality

The concept of research quality is also quite complicated to measure accurately. Most often, this construct is measured by the number of research publications or citations produced by faculty, or the amount of external research funding secured for research (Johnes 1988). Among these three methods, the number of publications perhaps best represents the quantity (and not quality) of research productivity, although there is a quality dimension to publication counts as well because research must pass through a peer review process (Toutkoushian et al. 2003).

In many contexts, citations can be viewed as the impact of an individual's research on others, and therefore they might capture some aspect of the quality of research produced. For example, Shanghai Jiao Tong University uses the number of most highly cited authors in their ranking indicators. Nonetheless, there is a large literature that discusses the pros and cons of using citations as a measure of productivity (see, for example, Toutkoushian 1994; Moed et al. 1985; Lindsey 1989).

Many rankers use the number of article published in internationally refereed journals, such as those monitored by the Institute of Scientific Information (ISI) or SCOPUS, as the measure of research productivity. However, many international journals may not be listed in ISI or SCOPUS because the indexing services often prefer English as the main language. Consequently, the international indexing service may understate the publication counts for faculty in many non-English speaking countries (Shin and Harman 2009; van der Wende and Westerheijden 2009). As a result, English speaking countries tend to dominate global rankings that rely on bibliometric measures of research productivity.

The number of citations is broadly accepted as a better indicator than publication counts of the quality of research. Many university rankings, especially global rankings, tend to emphasize citations as a measure of research productivity. However, there are controversies and issues on how to count citations. For example, some articles are cited by a journal with long reference lists while another article is cited in a journal with fewer references. In the scientometrics analysis, another issue that must be addressed is how to count the credit of co-authors. When an article is co-authored by professors in different institutions, the credit might be shared evenly by affiliated institutions or weighted differently by the contribution of each author, for example, first author, corresponding author, or second author. The current ranking surveys do not differentiate the different types of authorship in their ranking schemes, but the issue should be explored further so that rankings reflect the contribution of each author and hence each institution to research productivity. Currently, new methods to adjust these differences between journals and between disciplines are being developed by academic researchers (e.g., Leydesdorff and Opthof 2010; Moed 2010).

The extent of external research funding is often considered as a measure of quality as well as the quantity of research productivity. For example, external research funding is a popular indicator of productivity in the United States (e.g., Burke and Minassians 2002; Christal 1998). External research funding is based on rigorous peer review which is more in-depth and involves many stages, compared to journal publication decisions. Externally financed research funding received may have a quantity component to it because higher levels of funding may lead to greater amounts of research produced. It is important to take into account, though, that research funding varies significantly by discipline and that this measure best reflects the production process rather than the research output. In addition, research funding may also in part reflect quality because of the competition for research funding and the peer review process that applies to many externally funded research projects.

1.3.1.3 Measuring Service Quality

Service has received the least attention from academic researchers and rating agencies even though service is one of the three main functions of higher education institutions along with teaching and research. The omission is due in part to the difficulty in defining and measuring service quality. To faculty members, service may mean involvement in administrative activities on campus; however, in general, service represents the contributions of colleges and universities to society through means other than teaching and research. Higher education institutions provide many different types of service activities, such as working with local schools to improve the quality of education, helping government agencies make better decisions and policies, and assisting farmers with increasing their production of food. However, these contributions are also affected by the research and teaching that occurs within academia, which means that service quality is already partially reflected in teaching and research indicators. Because of the challenges involved in defining and measuring service quality, this remains an aspect of higher education that is underutilized in institutional rankings.

1.3.2 Considerations for Ranking Institutions

In calculating the total score, there are three major considerations that should be taken into account. First, a decision has to be made as to whether and how to control for the size of an institution in the rankings. If institutional size is not considered, then larger universities would tend to have advantages over equally productive but smaller institutions in the rankings game. If the goal of the rankings is to represent the relative productivity of an institution, then adjusting the rankings for institutional size would seem to be appropriate. An argument could also be made, however, that larger institutions should be ranked more highly than smaller institutions because in terms of sheer volume, they produce more outcomes in research, teaching, and service. Curiously, most institutional ranking schemes do not account for institutional size in

their rankings. An exception to this is the Shanghai Jiao Tong University ratings, where faculty size is included as an indicator. Other ranking schemes compare institutions on their output without considering the number of faculty or students as the denominator.

A second issue for institutional rankings is how to take into account the disciplinary differences across institutions. Some institutions are naturally more oriented toward the hard sciences, whereas others focus more on liberal arts. Because the publication and citation norms can vary across fields, programmatic differences across institutions may lead to differences in unadjusted productivity measures and hence rankings. This issue may be compounded by the fact that indexing services under-represent certain fields in their bibliometric counts, whereas bio and medical sciences are over-represented. Among the current rankings, the Leiden ranking is perhaps the only one that addresses the issue of disciplinary differences in its ranking scheme. This issue is another reason why some rankings focus exclusively on graduate programs in a single discipline rather than try to find an aggregate measure of productivity for a university.

A third methodological issue for rating agencies to address concerns the proper weightings of indicators in the overall rankings of institutions. Some indicators are typically weighted more highly than others. Among the three indicators (teaching, research, reputation, and internationalization) of the Times QS rankings, for example, institutional reputation is weighted (50%) more highly than other indicators such as scientific research productivity. The ranking status of individual institutions can vary depending on the selection of indicators to use and the weights attached to each. Unfortunately, there is little theoretical guidance that can be used to suggest the proper weights for indicators. The ambiguity of weights also leads to the development of new rankings which use different sets of indicators and weights.

1.4 Social Impacts of University Rankings

1.4.1 Impacts on Higher Education Systems

Ranking surveys, especially global rankings, have major impacts on higher education systems, higher education institutions, academics, and consumers (parents, students, employers). Schimank and Winnes (2000) conceptualized three types of higher education systems that combine research and teaching: the *French* model, the *Humboldtian* model, and the *post-Humboldtian* model. The French model is based on a division of labor between the research institute (e.g., academy) and the university, where the university is seen as the place for teaching but not for research. On the other hand, the Humboldtian model argues for combining research and teaching within the same university. In the post-Humboldtian model, American higher education institutions updated the Humboldtian model by establishing graduate schools that are independent from undergraduate programs (Schimank and Winnes 2000). The American model applied a division of labor at a university with the graduate school for research and the undergraduate education for teaching although faculty

members often perform tasks relating to both undergraduate and graduate education. The new model is based on empirical evidence that teaching is different from research (e.g., Marsh and Hattie 2002). The division of labor between teaching and research within faculty has also been applied in other countries including the UK, Netherlands, Australia, and Japan.

With the enormous interest in global rankings of colleges and universities, it is clear that they are contributing to systemic changes in many countries (Schimank and Winnes 2000). For example, France is initiating a merger of dual systems of research and teaching—research institute and university—into a single system by functionally merging the research institution and the university. A similar trend has been detected in ex-communist countries including Russia, China, and many Eastern European countries. Changes in higher education have also occurred in Germany, which views all universities as research universities regardless of their research performance across institutions. However, the German government has sought to establish a world-class research university by providing special research funding to high performing programs to help compete better with world-class universities in other countries. Similar changes have also been seen in the UK, although some of their universities already enjoyed high global reputations prior to the popularization of institutional rankings (Leistyte et al. 2009; Schimank and Winnes 2000).

1.4.2 Impacts on Universities and Students

For better or worse, university rankings have had major impacts on higher education institutions. In an attempt to increase their rankings, many universities began to focus more attention on publishing articles in journals that are included in bibliometric services such as ISI. Research universities have begun to transform faculty hiring and promotion policies to better reflect the new environment. For example, many research universities in Asian countries have begun to require a certain number of international journal publications as criteria in faculty hiring and/or promotion (Shin and Cummings 2010).

Another noticeable impact of ranking surveys on institutions is the growing emphasis on internationalization because this is a primary indicator in some ranking surveys. The Times QS rankings, for example, include student and faculty mobility as an indicator. Internationalization has been emphasized in many non-English speaking countries—especially in Asian countries where policymakers are sensitive to global rankings (Shin and Harman 2009). Some countries that use English as their primary language are successful in attracting international scholars and students. For example, institutions in Malaysia and Hong Kong often demonstrate outstanding performance in their ability to attract international scholars and students to their campuses.

As a way of encouraging internationalization, universities have also begun to emphasize English as their primary instructional language. For example, some Korean universities mandated students must take a certain number of courses in English.

In addition, universities are encouraged to hire international "star" faculty to enhance their reputation, establish international research networks, and enhance their global rankings. In addition, universities are enthusiastic about attracting international students to push up their rankings as well as to increase their financial resources.

Global rankings can also lead to changes in the academic culture of institutions. The traditionally conservative culture on many campuses has been giving way to a more innovative and market-oriented culture. Although many colleges still care quite a bit about reputation, and many rankings are based on an institution's reputation and not performance per se, policymakers and institutional leaders are increasingly seeking evidence of their excellence as well as their reputation. In addition, some universities have started to emphasize student satisfaction which has the potential to shift the culture in academia from the one oriented to professors to the one oriented to students (Usher and Medow 2009). As a result, today, universities are moving toward improving student satisfaction and their college experience because some rankers include student college experiences as an indicator. Nonetheless, reputation, prestige, and research are still major contributors to rankings of institutions and thus continue to play a large role in the activities of colleges and universities.

Finally, rankings are seen as a critical criterion taken into account by students when they are choosing a university (e.g., Bowman and Bastedo 2009; Drewes and Michael 2006). Researches based on the US and UK contexts have found that students from high-income and high-achieving family tend to rely heavily on rankings in their college choice (Dill and Soo 2005). Rankings are a particularly critical factor when international students choose to study abroad because they may find it difficult to visit an institution in another country prior to making a college decision. Students, especially from Asian countries are sensitive to rankings when they choose an international institution. In this respect, it would appear that the UK, Australia, and the United States are leaders in the rankings game because these countries have the highest share of international students in the world.

1.4.3 Side Effects of University Rankings

Although university rankings can have positive effects on institutions and students as discussed, there are many shortfalls in the practice of ranking institutions. Academic researchers, especially professors in social sciences and humanities, are quite concerned about the growth of rankings and their impact on academe. Dill (2009) summarizes the negative effects of rankings as a "highly costly, zero-sum game, in which most institutions as well as society will be the losers" (p. 102). In addition, other possible negative side effects of rankings might be characterized as institutional homogenization, distorting disciplinary balance, and leading institutions to change their focus and mission in response to rankings.

Although policymakers and institutional leaders emphasize the importance of institutional diversity, higher education institutions are considerably affected by the same set of ranking indicators. As van der Wende and Westerheijden (2009) argued,

"institutions act rationally and strategically, in effectively becoming what is being measured" (p. 77). Many university presidents consider getting a higher ranking position for their university as the main goal of their presidency, even though they know that the pursuit of rankings can have negative side effects.

Consequently, it has been asserted that universities are losing mission diversity because of the strong influence of rankings (van Der Wende and Westerheijden 2009; Dill 2009; Teichler 2009). The effect of rankings on institutions of higher education is conceptualized as homogenization or convergence of higher education institutions (Proulx 2009; Van Damme 2009). As Proulx (2009) argued, "the risk of isomorphism may represent the most important problem all the more" (p. 36). For example, with the emergence of global rankings, universities that seek to rise in the rankings may increase their emphasis on producing research, even when teaching has been their primary mission. Many teaching universities now take into account research performance when they hire their academic staff, which will lead to loose institutional missions in the near future. The changes are widely seen in Asian countries where the ranking of the university is critical in the social structure. Unfortunately, ranking surveys, especially global rankings, do not take into account institutional diversity and properly reward institutions for their production of teaching and service.

Secondly, as noted earlier, rankings do not consider disciplinary differences across institutions. Disciplines can differ in paradigms, preferred publication types, preferred research types (pure vs. applied), research methodology, time allocation between different types of academic activities (e.g., teaching, research, and services). Because of the strong influence of rankings on institutions, however, institutional leaders might weigh some disciplines more heavily than others when allocating resources setting, student quotas, and hiring faculty. In the rankings game, applied disciplines and bio-medical disciplines might benefit more than others because the faculty in applied and bio-medical disciplines, as well as hard sciences, tend to produce more publications than their peers in other fields.

A third concern is that in the ranking determination processes, higher education institutions are isolated from those doing the rankings (Locke et al. 2008). Rankers contact higher education institutions to collect data and to revise their ranking scheme, but the primary concerns of rankers, most of them being profit-generating media, are their customers (the readers of their ranking report). Higher education institutions often do not know how the raw data that they provide to rankers will be used, weighted, manipulated, and translated into rankings.

1.5 Future of University Rankings

Our discussion now moves to the types of rankings that can provide mutual benefits to rankers, higher education institutions, and their customers. The Berlin Principles provide implications for the future of rankings and how they may be improved. The Principles suggested that essential elements of university ranking should include

transparency, relevancy, and validity of comparative data. In addition, the following suggestions might contribute to the upgrading of current ranking systems:

- Current unified ranking systems should become *multiple ranking systems* to reflect different institutional missions, size, locations, etc.
- Ranker-centered systems should become *customer-centered systems* to satisfy readers' differing needs for rankings.
- Global ranking systems should become *regional ranking systems* to reflect regional characteristics, e.g., language, culture, etc.
- Institutional ranking systems should become *discipline-based ranking systems* in order to reflect disciplinary differences.

Among the current ranking systems, some rankers have been trying to reflect these dimensions to improve their ranking systems. The US News and World Report rankings provide various types of rankings according to institutional missions, regional locations, and sizes of institutions. These efforts help readers to select universities according to their primary interests, for example, top liberal arts university, best university in their region, etc. The Center for Higher Education (CHE) rankings (Germany) and the Maclean rankings (Canada) provide more accessible ranking services where customers can select indicators and apply different weightings based on their preferences. As van Der Wende and Westerheijden (2009) argued, "the real value of 'ranking' is not ranking, but matching" (p. 78).

Regional rankings of colleges and universities can attract interest because they can compare institutions with similar cultural boundaries. For example, the Times QS and Chosun Daily rankings (the latter a Korean daily newspaper) restricted their rankings to only Asian universities. The CHE rankings are also regional in scope, focusing on Germany and some European countries. The global ranking systems have also begun to provide disciplinary rankings as well as institutional rankings. For example, Shanghai Jiao Tong and Times QS provide disciplinary rankings.

In addition to the systemic changes in rankings, the possibility of combining current systems of measuring organizational effectiveness—such as measuring quality assurance and accountability—should be considered as well for institutional ranking schemes. One challenge is that the different systems require different indicators and they sometimes conflict with each other. Also, the different forms of indicators in these approaches require different forms of data, requiring higher education institutions to spend time managing and providing data for each system. The new mechanism for measuring organizational effectiveness should provide benchmarks for other institutions and provide institutional data from different angles to policymakers, institutional leaders, and students. The new systems should contribute to enhancing institutional performance as well as providing useful information to the consumers. The systemic changes and new directions of ranking survey might be combined as shown in Fig. 1.1.

Although the issues that we covered in this chapter are an introduction of the book, this chapter is a bridge to the theoretical and practical discussion of each chapter. University ranking is one dimension of the challenges that higher education is confronted with and a social phenomena associated with socio-economic

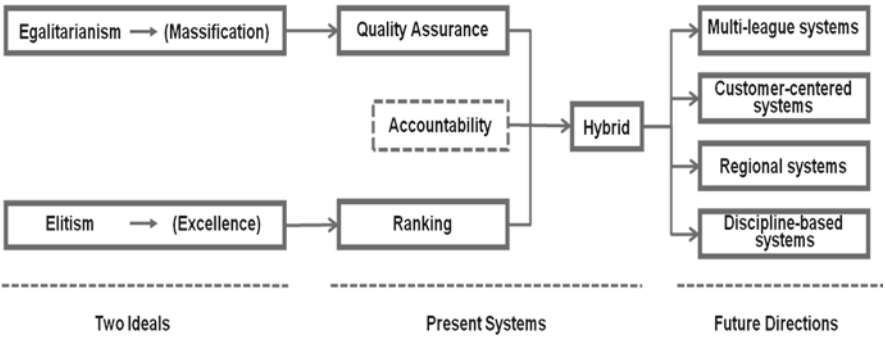

Fig. 1.1 Future directions for university rankings

contexts of contemporary higher education. We hope that readers find some insights on the past, present, and futures of rankings in this chapter.

References

Astin, A. (1993). *What matters in college: Four critical years revisited*. San Francisco: Jossey-Bass.

Borden, V., & Bottrill, K. (1994). *Performance indicators: History, definitions, and methods. New directions for institutional research, No. 82*. San Francisco: Jossey-Bass.

Bowman, N. A., & Bastedo, M. N. (2009). Getting on the front page: Organizational reputation, status signals, and the impact of U.S. News and World Report on student decisions. *Research in Higher Education, 50*(5), 415–436.

Brew, A., & Boud, D. (1995). Teaching and research: Establishing the vital link with learning. *Higher Education, 29*, 261–273.

Burke, J. C., & Minassians, H. (2002). Reporting indicators: What do they indicate? In J. Burke & H. Minassians (Eds.), Reporting higher education results: Missing links in the performance chain. *New Directions for Institutional Research, 116*, 33–58.

Cartter, A. (1966). *An assessment of quality in graduate education*. Washington, DC: American Council on Education.

Cave, M., Hanney, S., & Kogan, M. (1991). *The use of performance indicators in higher education: A critical analysis of developing practice* (2nd ed.). London: Jessica Kingsley.

Christal, M. E. (1998). *State survey on performance measures: 1996–97*. Denver, CO: State Higher Education Executive Officers/Education Commission of the States.

Conrad, C., & Blackburn, R. (1985). Correlates of departmental quality in regional colleges and universities. *American Education Research Journal, 22*, 279–295.

Dill, D. D. (2009). Convergence and diversity: The role and influence of university ranking. In M. K. Barbara & S. Bjorn (Eds.), *University rankings, diversity, and the new landscape of higher education* (pp. 97–116). Rotterdam: Sense Publishers.

Dill, D., & Soo, M. (2005). Academic quality, league tables, and public policy: A cross-national analysis of university rankings. *Higher Education, 49*, 495–533.

Drew, D., & Karpf, R. (1981). Ranking academic departments: Empirical findings and a theoretical perspective. *Research in Higher Education, 14*, 305–320.

Drewes, T., & Michael, C. (2006). How do students choose a university? Analysis of applications to universities in Ontario, Canada. *Research in Higher Education, 47*(7), 781–800.

Ehrenberg, R., & Hurst, P. (1996). The 1995 NRC ratings of doctoral programs: A hedonic model. *Change, 28*(3), 46–50.

Elton, L. (2001). Research and teaching: Conditions for a positive link. *Teaching in Higher Education, 6*(1), 43–56.

Feldman, K. A. (1989). Association between student ratings of specific instructional dimensions and student achievement: Refining and extending the synthesis of data from multisection validity studies. *Research in Higher Education, 30*, 583–645.

Hughes, R. (1925). *A study of the graduate schools of America*. Oxford, OH: Miami University Press, 3.

Hughes, R. (1934). Report on the committee on graduate instruction. *Educational Record, 15*, 192–234.

Johnes, G. (1988). Research performance indications in the university sector. *Higher Education Quarterly, 42*(1), 54–71.

Leistyte, L., Enders, J., & Boer, H. (2009). The balance between teaching and research in Dutch and English universities in the context of university governance reforms. *Higher Education, 57*(4), 509–531.

Leydesdorff, L., & Opthof, T. (2010). Scopus's source normalized impact per paper (SNIP) versus a journal impact factor based on fractional counting of citations. *Journal of the American Society for Information Science and Technology, 61*, 2365–2369.

Lindsey, D. (1989). Using citation counts as a measure of quality in science. *Scientometrics, 15*, 189–203.

Locke, W., Verbik, L., Richardson, J. T. E., & King, R. (2008) *Counting what is measured or measuring what counts? League tables and their impact on higher education institutions in England*, Report to HEFCE. Bristol: Higher Education Funding Council for England. http://www.hefce.ac.uk/pubs/hefce/2008/08_14/

Marsh, H. W. (1987). Students' evaluations of teaching: Dimensionality, reliability, validity, potential biases, and utility. *Journal of Educational Psychology, 76*, 707–754.

Marsh, H. W. (2007). Students' evaluations of university teaching: Dimensionality, reliability, validity, potential biases and usefulness. In R. P. Perry & J. C. Smart (Eds.), *The scholarship of teaching and learning in higher education: An evidence-based perspective* (pp. 319–384). New York: Springer.

Marsh, H. W., & Hattie, J. (2002). The relation between research productivity and teaching effectiveness: Complementary, antagonistic, or independent constructs? *The Journal of Higher Education, 73*(5), 603–641.

Moed, H. F. (2010). The source-normalized impact per paper (SNIP) is a valid and sophisticated indicator of journal citation impact. Available at: http://arxiv.org/abs/1005.4906

Moed, H., Burger, W., Frankfort, J., & van Raan, A. (1985). The use of bibliometric data for the measurement of university research performance. *Research Policy, 14*, 131–149.

Proulx, R. (2009). World university rankings: The need for a new paradigm. In M. K. Barbara & S. Bjorn (Eds.), *University rankings, diversity, and the new landscape of higher education* (pp. 35–46). Rotterdam: Sense Publishers.

Roose, K., & Andersen, C. (1970). *A rating of graduate programs*. Washington, DC: American Council on Higher Education.

Schimank, U., & Winnes, M. (2000). Beyond Humboldt? The relationship between teaching and research in European university systems. *Science & Public Policy, 27*(6), 397–408.

Shin, J. (2010). Impacts of performance-based accountability on institutional performance in the U.S. *Higher Education, 60*(1), 47.

Shin, J., & Cummings, W. (2010). Multilevel analysis of academic publishing across disciplines: Research preference, collaboration, and time on research. *Scientometrics*. doi:10.1007/s11192-010-0236-2.

Shin, J., & Harman, G. (2009). New challenges for higher education: Global and Asia-Pacific perspectives. *Asia Pacific Education Review, 10*(1), 1–13.

Simons, M., & Elen, J. (2007). The "research-teaching nexus" and "education through research": An exploration of ambivalences. *Studies in Higher Education, 32*(5), 617–631.

Teichler, U. (2009). Between over-diversification and over-homogenization: Five decades of search for a creative fabric of higher education. In M. K. Barbara & S. Bjorn (Eds.), *University rankings, diversity, and the new landscape of higher education* (pp. 155–182). Rotterdam: Sense Publishers.

Toutkoushian, R. (1994). Using citations to measure sex discrimination in faculty salaries. *The Review of Higher Education, 18*, 61–82.

Toutkoushian, R., Porter, S., Danielson, C., & Hollis, P. (2003). Using publication counts to measure an institution's research productivity. *Research in Higher Education, 44*, 121–148.

Usher, A., & Medow, J. (2009). A global survey of university rankings and league tables. In M. K. Barbara & S. Bjorn (Eds.), *University rankings, diversity, and the new landscape of higher education* (pp. 3–18). Rotterdam: Sense Publishers.

van Damme, D. (2009). The search for transparency: Convergence and diversity in the Bologna Process. In F. A. van Vught (Ed.), *Mapping the higher education landscape* (pp. 39–56). Dordrecht: Springer.

van der Wende, M., & Westerheijden, D. (2009). Rankings and classifications: The need for a multidimensional approach. In F. A. van Vught (Ed.), *Mapping the higher education landscape* (pp. 57–70). Dordrecht: Springer.

van Raan, A. F. J. (2007). Challenges in ranking universities. In J. Sadlak & L. Nian Cai (Eds.), *The world-class university and ranking: Aiming beyond status* (pp. 87–121). Bucharest: UNESCO-CEPES.

Webster, D. S., & Skinner, T. (1996). Rating Ph.D. programs: What the NRC report says…and doesn't say. *Change, 28*(3), 22–44. May/June 26.

Zheng, H. Y., & Stewart, A. A. (2002). Assessing the performance of public research universities using NSF/NCES data and data envelopment analysis technique. *AIR Professional File 83*. Tallahassee, FL: Association for Institutional Research.

Part I
Organizational Effectiveness, Quality, and Rankings

Chapter 2
Organizational Effectiveness and University Rankings

Jung Cheol Shin

2.1 Introduction

We pay attention to ranking and quality management systems because these mechanisms contribute to institutional quality and organizational effectiveness. Academics have believed that measuring organizational effectiveness in the public sector is much more difficult than in private corporations. The education field, especially higher education, has long been considered as an area where quality measures cannot be applied because professors designed their courses by themselves and they are recognized as having the highest specialty in the discipline areas. However, the perception has been changing with developments in academic theory and practice. For example, institutional leaders and theories have been developing measures of teaching quality by course evaluation and/or by student learning outcomes. Even van Vught (1995) argued that quality was an issue since university was established in the medieval ages. The French model was initiated to assure quality by external control (Catholic Church), while the English model was a self-governing model and the Italy model was by students who had the power of faculty hiring.

Today, higher education is no exception to the trends to assess organizational effectiveness. Government, research institutes, intermediate organizations, and the media are racing to develop quality mechanisms to enhance quality, to provide information to clients, or to expand their business markets. Recently, quality management has been applied in different types of higher education contexts. The growth of college rankings is a noticeable phenomenon worldwide, and every year, we notice that multiple numbers of rankings are released by new rankers, many of them being profit-generating news media. In addition, institutional leaders and board members are taking the ranking reports seriously because policymakers have begun to evaluate institutional leaders according to their ranking positions. The trends lead to much change on campus as institutional leaders react to the

J.C. Shin (✉)
Department of Education, Seoul National University, Seoul, South Korea
e-mail: jcs6205@snu.ac.kr

J.C. Shin et al. (eds.), *University Rankings*, The Changing Academy – The Changing Academic Profession in International Comparative Perspective 3,
DOI 10.1007/978-94-007-1116-7_2,

rankings by adopting managerial reforms to move their organization to a higher ranked position.

Ranking, however, does not guarantee that institutional quality is enhanced by moving toward a higher rank. In addition, higher education institutions spend their energy and resources to align with ranking indicators, although there is no clear evidence that ranking contributes to institutional effectiveness or institutional quality. This chapter discusses the theoretical grounds for organizational effectiveness and tries to link the ranking mechanism with organizational effectiveness perspectives which provide a more fundamental and broader view of ranking, its methodology, and its impacts on higher education.

2.2 Organizational Effectiveness and Quality

2.2.1 Approaches Toward Organizational Effectiveness

Organizational theorists have proposed theory on organizational effectiveness for many years, but organizational effectiveness is still a matter of controversy among academic researchers. Cameron (1981) summarized conventional approaches on organizational effectiveness in terms of four models—goal model, system resource model, process model, and participant satisfaction model.

- *Goal Model*: Effectiveness is measured by the extent to which the organization accomplishes its goals.
- *System Resource Model*: Effectiveness is measured by the extent to which the organization obtains needed resources.
- *Process Model*: Effectiveness is measured by organizational health, efficiency, and well-organized internal processes.
- *Participant Satisfaction Model*: Effectiveness is measured by the extent to which the need and expectations are met by the main constituencies.

(Cameron 1981: 25–26)

Each of these approaches represents a focus of one of the dimensions of organizational effectiveness and has strengths and shortfalls in defining and implementing organizational evaluations. For example, if we emphasize the goal model, we may ignore process or participant satisfaction. Defining organizational effectiveness in a perspective emphasizes only a dimension among multiple dimensions—goal, resource, processes, and participants—of organizational effectiveness. However, organizational emphases are changing even within the organization by changing organizational characteristics or by environmental changes. Quinn and Rohrbaugh (1983a,b) proposed a theoretical model which is known as Competing Value Model to provide a theoretical framework on dimensions of organizational effectiveness. In their competing value model, they classified organizational effectiveness on three dimensions—organizational focus, organizational structure, and organizational processes and goals.

Based on the three dimensions, they proposed four ideal types of organizational effectiveness: Open system model, rational goal model, internal process model, and human relations model (Quinn and Rohrbaugh 1983a,b). *Open system model* emphasizes flexibility and an external focus; *Rational goal model* emphasizes control and an external focus; *Internal process model* emphasizes control and an internal focus; and *Human relations model* emphasizes flexibility and an internal focus (Quinn and Rohrbaugh 1983a,b).

Since Quinn and Rohrbaugh proposed the competing value model, organizational researchers have applied multiple dimensions to explain organizational effectiveness. The *competing value model* contributes to understanding how organizational effectiveness is related to organizational characteristics and organizational culture. Under certain conditions, a specific model might be more effective than other types. For example, the open system model might be effective in the early stages of an organization, the rational goal model might be effective as it grows, the internal process model might be effective at maturity, and the human relations model might be effective when an organization is in turbulent situations (Quinn and Cameron 1983). In addition, organizational effectiveness might depend on organizational characteristics. In such organizations as educational institutions, the close relationships among administrators, teachers, and students are much more important than that with external constituencies. This is particularly true for higher education institutions.

In the recent turbulent challenges, all four of these dimensions might be required to enhance organizational effectiveness to comply with diverse *internal* and *external* demands through *stability* and *flexibility*. In their recent research on organizational effectiveness and culture in the US context, Smart (2003) concluded that "(the) improvement in the educational and managerial performance of college and universities is fundamentally tied to the development of campus cultures that reflect a healthy balance of the attributes reflected in the four cultural types…" (p. 699).

In the quality management field, the emphasis on these four dimensions is shifting from the human relations model toward the market model whether it is called as *academic capitalism* or *new public management* or as some other terms. This trend has been accelerated with the growing number of quality management schemes—ranking, performance-based accountability, or quality assurance. These three mechanisms have shared traits though they are slightly different in emphasis. Above all, they have developed evaluation indicators to measure organizational effectiveness and attempted to link their evaluation results with resources whether finance or reputation. Among these quality movements, ranking has had a huge impact on institutional management and academic society as well as on policymaking. However, this shift has not come without a cost.

2.2.2 From Effectiveness to Quality

Organizational research has focused on institutional quality as an indicator of organizational effectiveness. In organizational research, quality was considered as "the desired attributes of the outcomes produced by organizations"

(Cameron and Whetten 1996: 281). In their comprehensive literature review, however, Cameron and Whetten (1996) concluded that the concept of quality replaced the concept of organizational effectiveness in the mid-1980s. For example, they found that "quality" was the most frequently used term in academic journals and conferences in the early 1990s whereas "effectiveness" has disappeared from academic research.

What then is quality? Although in the recent literature quality is discussed more often than effectiveness, Winn and Cameron (1998) argued that the concept of quality is still under discussion and the literature has been focusing on "processes and procedures associated with reducing or preventing mistakes, controlling variations, or the dynamics associated with production of defect-free products or services" (p. 492). Winn (1996) summarized the definition of *quality* in higher education literature as: resource-based, content-based, outcome-based, value-added, constituency-based definitions, and productivity definition, and reputation definitions (for details, see Winn 1996 or Winn and Cameron 1998).

Compared to organizational effectiveness, Cameron and Whetten (1996) claimed that the concept of quality has three benefits. First, quality enables the integration of diverse perspectives of organizational effectiveness. Second, quality enables the integration of both tools (recourses and processes) and ends (outcomes) in producing better organizational performance. Third, quality has a comprehensive advantage because it covers a broader range of spectrums of effectiveness such as organizational culture.

Although the concept of quality is diverse and still under controversy, the creation of the Malcolm Baldrige National Quality Award by the US Department of Commerce in 1988 promotes the discussions on quality (Winn and Cameron 1998). The Malcolm Baldrige Award criteria have seven quality dimensions that explain *processes*, *procedures*, and *outcomes*: quality leadership, management of process quality, human resource development and management, strategic quality planning, quality information and analysis, customer focus and satisfaction, and quality and operational results. In Europe, European Quality Award was created in 1991. The quality framework includes eight components as its criteria of excellence: leadership, people, policy and strategy, partnerships and resources, processes, people results, customer results, society results, and key performance measures (Rosa and Amaral 2007). Although the dimensions of European award are slightly different from that of the USA, both have similarity in many aspects.

With the emergence of quality frameworks, each dimension of quality is not in competition with another as in the traditional effectiveness literature, but is integrated into a framework of total quality for the organization. Nevertheless, the quality management frameworks are developed to apply in private sector; thus, there have been controversies on the relevance of quality management tool in public sector, especially, education areas. In higher education research, some studies have applied the frame in higher education institutions. For example, Winn and Cameron (1998) applied Baldrige criteria in the US university and found that these dimensions are applicable in the university. They found that leadership indirectly affects outcomes (customer focus and satisfaction, and quality and operational results)

through mediating factors (management of process quality, human resource development and management, strategic quality planning, quality information and analysis). Also, Rosa and Amaral (2007) tested a possibility of applying European Quality Award framework in a Portugal context. The pioneering efforts have contributed to attract institutional leaders' attention to institutional effectiveness in higher education contexts.

2.3 Measuring Organizational Effectiveness

Whether we focus on effectiveness or quality, the practical issue is how to measure these. Because most discussions on quality and effectiveness have been developed through theoretical discussions rather than empirical data, the measure of effectiveness is still controversial. As Cho (2007) argued, organization cannot improve its effectiveness if we do not have clear definitions on what effectiveness is. Considerable research has been conducted by Cameron (1978, 1981), including follow-up studies to bridge the gap between theory and empirical data. The next section focuses on Cameron's dimensions of effectiveness and his follow-up studies.

2.3.1 Measure of Organizational Effectiveness

Cameron (1978) proposed a model of organizational effectiveness, and the model has been widely applied in higher education research. He identified nine dimensions of organizational effectiveness in his study on higher education institutions in the USA—four related to students and five to staff and institutions. The nine dimensions are: student's educational satisfaction, student's academic development, career development, personal development, faculty and administrator employment satisfaction, professional development and quality of faculty, system openness and community interaction, ability to acquire resources, and organizational health. Many researchers have confirmed Cameron's dimensions in different study settings and in different research focus, and even different cultural contexts.

For example, Cameron's instrument has been applied and confirmed in the UK, Australia, and Hong Kong (Kwan and Walker 2003; Lysons and Hatherly 1992; Lysons et al. 1998). Other researchers have used Cameron's instrument to explore the link between organizational effectiveness and organizational culture (e.g., Cameron and Ettington 1988; Smart and St. John 1996), leadership and management (Koh et al. 1995; Williams et al. 1992), and governance and organizational structure (e.g., Kushner and Poole 1996; Schmid 1992). Interestingly, however, these variables were not consistently identified in different research settings although these dimensions of organizational effectiveness contribute to the measurement of organizational effectiveness.

Recently, Kwan and Walker (2003) applied Cameron's instrument in Hong Kong and revised the dimensions of the model. They excluded one dimension (organizational health) of Cameron's nine, and found seven dimensions in their data analysis. They found that student educational satisfaction is combined with student personnel development; integrated ability to acquire resources with professional development and quality of faculty; but they split community interaction and system openness into *system openness* and *community interaction*. Student satisfaction and their personnel development represent student campus life or preparation for more long-term life, and thus both have commonality. External resources were also closely related with institutional reputation which is mainly influenced by faculty performance; thus, both *acquire resources* and *professional development and faculty quality* have commonality. In general, community interaction represents faculty participation in community services, while system openness represents how to satisfy the community's demand on the education program. Both represent different dimensions of organizational effectiveness.

2.3.2 Considerations in Measuring Organizational Effectiveness

In assessing organizational effectiveness, we usually apply an *overall* rating of effectiveness because people tend to compare one organization with others in terms of overall score. The reputation ranking is the overall rating of institutional effectiveness. Cameron and Whetten (1996: 275) proposed seven guidelines for measuring organizational effectiveness: time frame, level of analysis, perspective of effectiveness, domain of activity, purpose for judging effectiveness, types of data being used, and reference of judging effectiveness.

The meaning of these six criteria is clear if we compare organizational effectiveness between different types of quality mechanisms. As an example, we compare ranking and quality assurance because quality assurance is at the other extreme from ranking in terms of quality improvement.

- *Time frame*: Quality assurance is a longitudinal assessment, while ranking is an annual event. Quality assurance has more influence on longitudinal changes, while ranking focuses on short-term change. In reality, institutional changes are not short-term, although, ranking shows how institutional rankings have shifted compared with the previous year.
- *Level of analysis*: Quality assurance focuses on a multilevel structure of institutional effectiveness, while ranking mainly focuses on overall ratings. Quality assurance considers program quality as well as institutional quality overall. In reality, academics are more interested in their program rankings than overall institutional rankings. Thus, institutional administrators are interested in an overall ranking, while academics are interested in program-level evaluation.
- *Main constituency*: Quality assurance emphasizes perspectives from higher education institutions, while ranking focuses on perspectives from media, students, parents, and the general public.

- *Domain of activity*: Quality assurance allows for a higher weighting on teaching quality, while ranking, especially international ranking, focuses on research productivity. However, the difference is not generally applicable in domestic rankings because many domestic rankings weight teaching, too.
- *Purpose of evaluation*: Quality assurance seeks to enhance institutional quality, while ranking is mainly interested in the relative positions between comparable institutions. Thus, quality assurance is more interested in benchmarks and best practices, while ranking is interested only in rank position.
- *Type of data*: Quality assurance is based on qualitative as well as quantitative data, while ranking is mainly based on quantitative data.
- *Reference of judgment*: Quality assurance uses benchmark in many cases, while ranking does not have benchmarks. Quality assurance is therefore able to enhance the quality of institutions, while ranking may or may not.

These comparisons between quality assurance and ranking are in general terms. Clearly different types of quality assurance and ranking might provide different comparisons. For example, rankings by the Centre for Higher Education and Development (CHE) in Germany and Maclean are customer-oriented rankings. These rankings do not provide ordinal ranking but focus on providing institutional information to the public. They have commonality with quality assurance and show deviance from the other types of rankings. Considering the features of CHE and Maclean, both have more in common with quality assurance, while other rankings have quite heterogeneous characteristics from quality assurance.

2.4 University Rankings as a Measure of Organizational Effectiveness

In higher education, various mechanisms for quality measurement have been developed. These are quality assurance, accountability, ranking, academic program review, follow-up studies, total quality management, etc. (Bogue and Hall 2003). Some of them have been developed by the public sector, while others have been developed by private sectors. The relatively well-known mechanisms are quality assurance, accountability, and ranking. In this chapter, we will overview these three mechanisms and will discuss in detail in Chap. 3.

2.4.1 Ranking, Quality Assurance, and Accountability

Although these three mechanisms have much in common because they provide information to the public and enhance institutional quality, they differ in their goals, method of evaluation, publishing of results, and policy links. The differences are caused by their goals and by their target customers. For example, the primary goal of quality assurance is enhancing institutional quality as defined by institutional

mission, while ranking focuses on ranking order and accountability in order for the legitimate use of public taxes. Details of these three types are explained below.

- *Main stakeholders*: Government, higher education institutions, and quality agencies are the main stakeholders for quality assurance; news media for rankings; and government and funding agencies for accountability.
- *Forms of actions*: Accreditation, quality assessment, and quality audit are the main forms of quality actions for quality assurance; rank order for rankings; and performance-based funding/budgeting, and performance reporting for accountability.
- *Indicators*: Teaching has priority in quality assurance; research in ranking especially worldwide ranking; and teaching and service in accountability.
- *Data sources*: Peer review, nationwide data, and survey data are the sources for quality assurance; nationwide data and survey data for rankings; and nationwide (statewide data in the US contexts) data for accountability.
- *Linking with government policy*: Quality assurance is closely linked with an institution's legal status, financial aids, and funding policy; ranking does not have direct linking with government policy; and accountability is directly or indirectly linked with government policy.
- *Main customers*: HEIs and government are the main stakeholders for quality assurance; parents and students, and HEIs for rankings; and government for accountability.

The summary of comparisons between quality assurance, ranking, and accountability is provided in Table 2.1.

Table 2.1 Comparisons between quality assurance, ranking, and accountability

Characteristics	Quality assurance	Ranking	Accountability
Goals	Enhancing quality	Information providing	Financial accountability
Stakeholder	Government/HEIs/ agency	Media/research institute	Government/funding agency
Actions	Accreditation Quality assessment Quality audit Program review Licensure	Ranking by institution Ranking by region or disciplines	Performance reporting Performance-funding/ budgeting
Indicators	Teaching/research/ service	Research/teaching/ reputation/ internationalization	Teaching/research/ service
Data sources	Nationwide data Peer review/survey	Nationwide data Peer review/survey	Nationwide data
Linking with government policy	Institution's legal status Financial aids Research funding Operational funding	Not linking Some developing countries link with policy	Linking or not linking with funding
Customers	HEIs, government	Parents, students, HEIs, enterprise, government	Government

2.4.2 Landscapes of Rankings Worldwide

University rankings have been used in the USA since 1925 when Raymond Hughes reported reputational ranking of US graduate programs. Since then, scholars have conducted rankings in similar ways to Professor Hughes. For example, Hayward Kenisoon in 1959, Allan Cartter in 1966, Roose and Anderson in 1970, Mary Jo Clark, Rodney Hartnett, and Leonard Baird in 1976, and the National Academy of Science in 1982 (Bogue and Hall 2003). These rankings are based on reputations from peers and focus on graduate programs. The turnover on rankings was started in 1983 when the *US News and World Report* published its first ranking report, "America's Best Colleges." The US News ranking report is distinct from other rankings on two points. First, the *US News* focused on undergraduate courses whereas previous rankings focused on graduate programs. Second, the *US News* rankings were initiated by the news media for the business purposes.

Other news media began to enter the ranking market worldwide during the 1990s. For example, the *Guardian* in the UK in 1999, *Maclean* in Canada in 1990, *Jungang Daily* in South Korea in 1997, etc. In addition, worldwide ranking reports have been released in the 2000s. For example, *Shanghai Jiao Tung* (*SJTU*) reported world-class university rankings in 2003 followed by *The Times* in 2004, *Webometric* in 2004, and *Taiwan Higher Education and Accreditation Council* in 2007. According to Usher and Medow (2009), there were 26 rankings in the world in 2007; however, these did not include some of rankings in Asian countries, e.g., *Jungang Daily* in Korea, *Asahi News* in Japan, etc.

There have been two challenges to media-led rankings. Their first challenge is in relation to data reliability and the validity of measures. Secondly, ranking does not provide information on how to improve institutional quality because it simply provides ranking information. Academic research institutes (e.g., Shanghai Jiao Tong, Melbourne Institute, Center for Higher Education, etc.) provide better more valid, and more reliable rankings than the commercially initiated ones and contribute to improving quality rather than the ordinal rating itself.

2.4.3 Measures of Rankings

The primary goal of rankings is to provide information to students and parents for college choice, as well as in relation to the quality of institutions. We analyzed goal statements of rankings from four worldwide and five nationwide surveys. We found that four of them provided information both for college choice and for quality of the organization, while two of them (Shanghai Jiao Tong and Higher Education Evaluation and Accreditation Council) provided only quality information. Interestingly, many rankings are seeking to provide information on institutional quality to the public; yet, the indicators they use are opposite to their stated goals. For example, five of them include reputation as well as teaching and research indicators.

Table 2.2 Comparisons of ranking by goals and indicators

Rankings		Goals		Indicator weights (%)			
		Choice	Quality	Teaching	Research	Reputation	Internationalization
Worldwide	Times	0	0	20	20	50	10
	US News	0	0	20	20	50	10
	SJTU	×	0	30	70	–	–
	HEEAC	×	0	–	100	–	–
Nationwide	US News	0	0	70	–	30	–
	Maclean	0	×	78	–	22	–
	CHE	0	0	–	–	–	–
	Netbig	0	0	63	22	15	–
	Jungang	0	0	43.7	23.8	15	17.5

Notes:

(a) If the ranking officially announces the main goal is to help students' college choice, it is coded as "choice"; if the ranking announces the main goals is to provide information for institutional quality improvement, it is coded as "Quality"

(b) *HEEAC* is worldwide ranking by Higher Education Evaluation and Accreditation Council in Taiwan

(c) *CHE* is a ranking by the Centre for Higher Education and Development in Germany

(d) *Netbig* is ranking for colleges and universities in China

Reputation is a perception of those surveyed rather than a measure of institutional performance or quality. This feature implies that rankings, that are weighted heavily on reputation (e.g., *The Times QS*, the *US News* world ranking), are based on perceptions of those surveyed and do not reflect institutional quality (Table 2.2).

Another feature of ranking indicators is the emphasis on research performance. Rankings, especially worldwide rankings, emphasize research performance. This may be related to the emergence of globalization and the knowledge-based economy. As policymakers recognize higher education institutions as the center of global competition, rankers began to purposely focus on the quality and quantity of faculty research. For example, SJTU ranking focuses on research because their goal is "to find out the gap between Chinese universities and world-class universities" (Liu et al. 2004).

In reality, research performance is one of the rare criteria which enable the comparison of higher education institutions worldwide because there are comparable data, e.g., Institute of Scientific Information (ISI), SCOPUS, etc. However, many other criteria are complicated by socio-economic contexts. Even the numbers of full-time faculty differ depending on national contexts. We can compare faculty-student ratio in the US contexts, but may not compare faculty-student ratio across countries because each country use different terms of full-time faculty. Finally, the top ranked institutions in worldwide rankings are all research-focused universities, which suggest that research performance might be a better indicator than others such as graduates' employment rate, their educational satisfactions, etc.

Compared to worldwide rankings, domestic ranking surveys do not pay much attention to research productivity but emphasize the reputation of each institution. This makes sense because many academics already know the performance of

their peers in their own country and their peers in their competing institutions. Also, reputations among peers reflect institutional quality better than quantified measures; or even reputation has high correlations with performance measures (Williams and Dyke 2008). This chapter will not discuss ranking measures in detail. We will discuss ranking methodology issues in Part II.

The next question to consider is "are ranking measures related to measures of quality or organizational effectiveness?" When considering ranking as a way of measuring institutional effectiveness or performance, it should reflect dimensions of organizational effective or quality. For simplicity, we focus on the measures of organizational effectiveness rather than quality because organization theory developed measures of organizational effectiveness much earlier, and organizational researchers have applied relatively consistent measures of effectiveness in different research settings. To develop our idea on how the ranking measures are related to effectiveness measures, a comparison table is provided in Table 2.3. In the table, the dimension of organizational effectiveness is based on Kwan and Walker (2003), which is the revised version of Cameron's study in 1978. Kwan and Walker's study is the most recent study, which applied Cameron's study out of the US contexts; so, more makes sense to the rest of the world.

Among the seven dimensions of organizational effectiveness, five dimensions have been included in many domestic or international rankings. Exceptions are faculty employment satisfaction and community interaction dimensions. Rankings, whether domestic or international, do not pay much attention to faculty satisfaction, while employee's job satisfaction is a critical factor for other organizations. Here, a question emerges.

Table 2.3 Dimensions of organizational effectiveness and ranking indicator

Dimension	Areas of measure	Method of data collection	Inclusion in ranking indicator
Student educational satisfaction and personal development	Campus life	Graduate survey/ engagement survey	Yes (domestic ranking)
Student academic development	Teaching	Class evaluation/graduate survey/standardized test	Yes (domestic ranking)
Student career development	Teaching	Follow-up survey/employer satisfaction survey	Yes (domestic ranking)
Faculty employment satisfaction	Campus life	Faculty satisfaction survey	No
Faculty professional development and institutional ability to acquire resources	Research	Research productivity (publication, citation)	Yes (worldwide and domestic ranking)
System openness	Employer satisfaction	Employer satisfaction survey	Yes (worldwide and domestic ranking)
Community interaction	Service	Survey	No

Why are rankers not interested in faculty job satisfaction? In some respects academics are not employees but are self-employed, although the university hired them. Historically, faculty has been independent from state or institutional control to some extent. This may be true of prestigious institutions where faculty have greater academic freedom but might not be true of other more recently established institutions, where faculty have heavier workloads and are under-resourced. However, ranking was designed to lead to competition among academics and to enhance institutional quality. Rankers, especially media-led rankers, are not much interested in the quality of academic life; rather, commercial rankers are more interested in how to attract audiences and thus to generate benefits from selling rankings.

In addition, rankers have not paid much attention to community interaction (faculty participation in community activity). In faculty evaluation, a growing number of institutions tend to see community interaction as an indicator of faculty performance (O'Meara 2002, 2005). However, rankers may not find it easy to include community activity as an indicator of ranking because communities where higher education institutions are based have different types of demands on the institutions. If ranking is about the comparisons between similar institutions in its mission, rankers may include community interaction as an indicator, but if it is not, they are unlikely to include it.

In summary, most dimensions of organizational effectiveness have been reflected in ranking indicators, whether it is a worldwide or domestic ranking. Two dimensions that might not match up with rankings were not included in the rankings. Therefore, we conclude that rankings evaluate organizational effectiveness to some extent, though each dimension is represented by a limited numbers of indicators.

2.5 Impacts of Rankings on University Effectiveness

We are moving toward a new question: Do university rankings contribute to organizational effectiveness? It is a more critical question than simply asking whether rankings reflect the dimensions of organizational effectiveness because the question is about the legitimacy of rankings. If they do not contribute to organizational effectiveness, we may no longer need ranking. Alternatively, we may have to simply provide institutional information to the public whether about institutional quality or financial resources, student academic preparation, curriculum, and so on.

There are few empirical studies on the impact of rankings on institutional effectiveness, although academics have argued about the negative effects of ranking on institutional mission diversity, management, and faculty work-life (e.g., Marginson and Van der Wende 2007; Teichler 2008). Many academic researchers have focused on methodological issues related to rankings rather than on their impact on institutional effectiveness. The majority of impact studies have focused on the impact of rankings on a student's college choice, donation, etc. (Bowman and Bastedo 2009; Drewes and Michael 2006; Hazelkorn 2008; Merddith 2004).

To the best of my knowledge, however, there is no clear evidence that ranking contributes to quality or organizational effectiveness. Some studies have found that ranking has an impact on student's college choice, but there is no clear causal relationship between a ranking report and its impacts on institutional quality.

There have also been some studies done on the impact of quality assurance and accountability on institutional performance (e.g., Brennan and Shah 2000; Shin 2010). Brennan and Shah (2000) conducted case studies on 29 institutions from 14 countries and found that the quality assurance framework in each country has impacts on institutional quality, although the impacts are different depending on the mechanism that each country or each institution is based on. Volkwein and Tandberg (2008) and Shin (2010) analyzed the impacts of performance-based accountability on institutional performance in the US context and concluded that accountability did not contribute to institutional performance. In their international comparisons, Huisman and Currie (2004) and Himanen et al. (2009) found that the performance-based accountability did not contribute to institutional performance in the countries they studied.

Notwithstanding the fact that academics have found that ranking has many negative side effects on higher education institutions, institutional leaders and policymakers have not paid attention to these issues. If ranking does not contribute to institutional quality, but simply provides information for college choice, it may lose its legitimacy. Alternatively, government agencies (e.g., education statistics providers), or university associations might provide more comprehensive and reliable information to students and parents, as well as for academic researchers.

What changes does ranking bring to higher education institutions? A consideration of the impact of ranking on institutional quality leads to the question of whether we need ranking given that it may have possible negative effects. As higher education scholars, our task is to scrutinize theoretical and practical issues such as whether rankings have any impact on institutional quality—teaching, research, and service.

2.6 Concluding Remarks

We discussed university rankings from the viewpoint of organizational effectiveness. In this chapter, we compared how the dimensions of organizational effectiveness are matched with those measured by ranking. We showed that the dimensions that ranking measures are quite similar to that of organizational effectiveness, although some dimensions are not included in ranking. We also examined whether ranking affects institutional quality. There is no clear evidence that ranking contributes to institutional quality, while ranking appears to have many negative effects on higher education institutions. The issue is how to minimize the problems that ranking brings to higher education if we still need ranking. This is one of the main purposes of our exploration of these issues in this book.

The contribution of ranking might be similar to that of quality assurance if we develop better ranking systems providing reliable and comprehensive data to

students and policymakers. Good ranking will include qualitative and quantitative indicators, and reflect customer satisfaction as well as expert evaluation in its judgment. It will also contribute to institutional quality. Because ranking, accountability, and quality assurance have much in common, they may eventually converge on a single quality mechanism. They may share indicators and data collection procedures. However, although they share its indicators and data collection, the judgment criteria might be divergent depending on the main goals of each approach.

Finally, we looked at the impacts of ranking on institutional quality. Studying ranking impacts will contribute to further discussions on ranking and its related policy issues. In addition, economic issues should not be underestimated in any study of rankings. What is the cost of releasing a ranking report? The cost paid by the media is only a fraction of the real costs. Higher education institutions and students pay most of costs accompanied by ranking release. Meanwhile, most of benefits might be enjoyed by rankers, profit generating media. These dimensions are not addressed enough by higher education researchers. In the future, therefore, critical research topic is *the economics of ranking* in ranking study.

References

Bogue, G., & Hall, K. B. (2003). *Quality and accountability in higher education: Improving policy, enhancing performance*. Westport: Praeger Publishers.

Bowman, N. A., & Bastedo, M. N. (2009). Getting on the front page: Organizational reputation, status signals, and the impact of U.S. News and World Report on student decisions. *Research in Higher Education, 50*(5), 415–436.

Brennan, J., & Shah, T. (2000). *Managing quality in higher education: An international perspective on institutional assessment and change*. Buckingham: Organization for Economic Co-operation and Development.

Cameron, K. S. (1978). Measuring organizational effectiveness in institutions of higher education. *Administrative Science Quarterly, 23*, 604–632.

Cameron, K. S. (1981). Domains of organizational effectiveness in colleges and universities. *Academy of Management Journal, 24*(1), 25–47.

Cameron, K. S., & Ettington, D. R. (1988). The conceptual foundations of organizational culture. In S. John (Ed.), *Higher education: Handbook of theory and research* (Vol. 4, pp. 356–396). New York: Agathon Press.

Cameron, K. S., & Whetten, D. A. (1996). Organizational effectiveness and quality: The second generation. In S. John (Ed.), *Higher education: Hand book of theory and research* (Vol. XI, pp. 265–306). New York: Agathon Press.

Cho, S. M. (2007). Assessing organizational effectiveness in human service organizations: An empirical review of conceptualization and determinants. *Journal of Social Service Research, 33*(3), 31–45.

Drewes, T., & Michael, C. (2006). How do students choose a university? Analysis of applications to universities in Ontario, Canada. *Research in Higher Education, 47*(7), 781–800.

Hazelkorn, E. (2008). Learning to live with league tables and ranking: The experience of institutional leaders. *Higher Education Policy, 21*, 193–215.

Himanen, L., Auranen, O., Puuska, H., & Nieminen, M. (2009). Influence of research funding and science policy on university research performance: A comparison of five countries. *Science and Public Policy, 36*(6), 419–430.

Huisman, J., & Currie, J. (2004). Accountability in higher education: Bridge over troubled water? *Higher Education, 48*, 529–551.

Koh, W., Steers, R., & Terborg, J. (1995). The effects of transformational leadership on teacher attitudes and student performance in Singapore. *Journal of Organizational Behavior, 16*, 319–333.

Kushner, R., & Poole, P. (1996). Exploring structure-effectiveness relationships in nonprofit arts organizations. *Nonprofit Management & Leadership, 7*, 119–136.

Kwan, P., & Walker, A. (2003). Positing organizational effectiveness as a second-order construct in Hong Kong higher education institutions. *Research in Higher Education, 44*(6), 705–726.

Liu, N. C., Cheng, Y., & Liu, L. (2004). Academic ranking of world universities using scientometrics – A comment to the "Fatal Attraction". *Scientometrics, 64*(1), 101–109.

Lysons, A., & Hatherly, D. (1992). Cameron's dimensions of effectiveness in higher education in the U.K.: A cross-cultural comparison. *Higher Education, 23*, 221–230.

Lysons, A., Hatherly, D., & Michell, D. A. (1998). Comparisons of measures of organizational effectiveness in U.K.: Higher education. *Higher Education, 36*, 1–19.

Marginson, S., & Van der Wende, M. (2007). To rank or to be ranked: The impacts of global rankings in higher education. *Journal of Studies in International Education, 11*(3/4), 306–329.

Merddith, M. (2004). Why do universities compete in the ratings game? An empirical analysis of the effects of the U.S. News and World Report college rankings. *Research in Higher Education, 45*(5), 443–461.

O'Meara, K. (2002). Uncovering the values in faculty evaluation of service as scholarship. *Review of Higher Education, 26*(1), 57–80.

O'Meara, K. (2005). Encouraging multiple forms of scholarship in faculty reward systems: Does it make a difference? *Research in Higher Education, 46*(5), 479–510.

Quinn, R. E., & Cameron, K. (1983). Organizational life cycle and shifting criteria of effectiveness: Some preliminary evidence. *Management Science, 29*(1), 33–51.

Quinn, R. E., & Rohrbaugh, J. (1983a). A competing values approach to organizational effectiveness. *Pubic Productivity Review, 5*, 122–140.

Quinn, R. E., & Rohrbaugh, J. (1983b). A special model of effectiveness criteria: Toward a competing values approach to organizational analysis. *Measurement Science, 29*, 363–377.

Rosa, M. J., & Amaral, A. (2007). A self assessment of higher education institutions from the perspective of the EFEM excellence model. In D. F. Westerheijden, B. Stansaker, & M. J. Rosa (Eds.), *Quality assurance in higher education: Trends in regulation, translation and transformation* (pp. 181–208). Dordrecht: Springer.

Schmid, H. (1992). Relationship between decentralized authority and other structural properties in human service organizations: Implications for service effectiveness. *Administration in Social Work, 16*(1), 25–39.

Shin, J. C. (2010). Impacts of performance-based accountability on institutional performance in the U.S. *Higher Education, 60*(1), 47–68.

Smart, J. C. (2003). Organizational effectiveness of 2-year colleges: The centrality of cultural and leadership complexity. *Research in Higher Education, 44*(6), 673–703.

Smart, J. C., & St. John, E. P. (1996). Organizational culture and effectiveness in higher education: A test of the "culture type" and "strong culture" hypotheses. *Educational Evaluation and Policy Analysis, 18*(3), 219–241.

Teichler, U. (2008). Diversification? Trends and explanations of the shape and size of higher education. *Higher Education, 56*(3), 349–379.

Usher, A., & Medow, J. (2009). A global survey of university rankings and league tables. In M. K. Barbara & S. Bjorn (Eds.), *University rankings, diversity, and the new landscape of higher education* (pp. 3–18). Rotterdam: Sense Publishers.

Van Vught, F. (1995). The new contexts for academic quality. In D. Dill & B. Sporn (Eds.), *Emerging patterns of social demand and university reform: Through a glass darkly* (pp. 194–211). Oxford: Pergamon.

Volkwein, J. F., & Tandberg, D. A. (2008). Measuring up: Examining, the connections among state structural characteristics, regulatory practices, and performance. *Research in Higher Education, 49*, 180–197.

Williams, R., & Dyke, N. V. (2008). Reputation and reality: Ranking major disciplines in Australian universities. *Higher Education, 56*, 1–28.

Williams, M., Podsakoff, P., & Huber, V. (1992). Effects of group-level and individual-level variation in leader behaviors on subordinate attitudes and performance. *Journal of Occupational and Organizational Psychology, 65*, 115–129.

Winn, B. (1996). *Organizational quality in higher education: An examination of the Baldrige framework in the university work environment*. Unpublished doctoral dissertation, Center for the Study of Higher and Postsecondary Education, University of Michigan, Ann Arbor.

Winn, B. A., & Cameron, K. S. (1998). Organizational quality: An examination of the Malcolm Baldrige national quality framework. *Research in Higher Education, 39*(5), 491–512.

Chapter 3
Competitors of Rankings: New Directions in Quality Assurance and Accountability

Grant Harman

3.1 Introduction

This chapter is concerned with the development of modern higher education quality assurance systems and how in recent years these systems have interacted with both national and global institutional ranking. Among various perspectives employed is that of organisational effectiveness which provides a theoretical orientation for the volume. Quality assurance, accountability and rankings can be viewed as different forms of assessment of organisational effectiveness. The chapter focuses particularly on recent fast-moving political and economic currents that are forcing substantial changes to existing quality assurance and accountability systems. These currents are being driven by a combination of stakeholder and government concerns about academic standards, global competition for education services and the impact of ranking systems. These recent changes are set within the broader context of the development of modern quality assurance since the 1980s. Unlike many discussions of quality assurance, this chapter sees assessments of research quality and impact as key elements of quality assurance. Accountability, however, is considered only in the narrow context as one purpose of quality assurance.

Over the past two to three decades, quality assurance has gained a central place on higher education policy agendas internationally. In many countries, national quality assurance systems have had significant impacts, especially on the governance and internal management of higher education, and relationships between higher education and government and between higher education and external stakeholders. In some countries, quality assurance systems are well institutionalised and well accepted, while in others they are still being contested, or are in an early stage of development. Internally, within higher education institutions, quality assurance has impacted particularly on how institutions plan future directions, develop and

G. Harman (✉)
Centre for Higher Education Management and Policy, University of New England, Armidale, NSW, Australia
e-mail: gharman@une.edu.du

J.C. Shin et al. (eds.), *University Rankings*, The Changing Academy – The Changing Academic Profession in International Comparative Perspective 3,
DOI 10.1007/978-94-007-1116-7_3,

review their courses and academic units, handle student assessment, decide what information about academic standards and achievements is conveyed to students and the wider community, and allocate research resources between different academic units. With the ongoing transition towards knowledge-driven economies, education has taken on greater importance, highlighting the need for high rates of student participation and for post-school education provision to be of undoubted quality. Quality courses are essential if graduates are to effectively participate in the labour market and society at large.

More recently, new policy currents in which rankings are of key importance, are heightening government concerns about quality, particularly the quality of graduates and research outputs (Salmi 2008), and are having a major impact on quality assurance systems. Various official reports and studies (Tremblay 2009; Institute for Higher Education Policy 2009; Stella and Woodhouse 2006) point to the substantial impact of rankings on quality assurance systems, quality assurance agencies and higher education more generally. While national quality assurance agencies have generally been critical of ranking systems, pointing to doubtful methodologies, arbitrariness in choice of indicators, lack of transparency and bias towards larger and English speaking research-intensive institutions, both global and national rankings appear to be satisfying public demand for information and are influencing prospective students' choices. Rankings have become new forms of 'super' quality assessments that have considerable attraction for Ministers and the public, especially since quality is expressed in simple numerical scores or league tables. Rankings have thus 'cemented the notion of a world university market' (Marginson and van der Wende 2006) in which higher education institutions are measured according to their relative standing on a global scale. This, in turn, is prompting governments and universities to take a much keener interest in both the global indicators of quality, academic standards and student outcomes, while governments are seeking much greater direct control over quality assurance agencies. The relationships between quality assurance and global rankings are complex. Not only do global rankings employ data and methodologies that in many cases were developed originally for quality assurance purposes, the ranking systems are tending to over-shadow quality assurance assessments (Usher and Salbvino 2007).

Quality assurance refers to national and institutional systems designed to assess and improve the quality of teaching and research, and provide relevant information to key stakeholders on academic standards and employment of graduates. It is important, however, to recognise that different scholars define quality assurance in different ways, reflecting different conceptions of what is quality in higher education, and how it can be best assessed and developed. Harman and Meek (2000), for example, define quality assurance as specialised management and assessment procedures adopted by higher education institutions or systems for the purpose of monitoring and enhancing academic performance, safeguarding academic standards and generating information on outputs and standards for stakeholders. This definition highlights the need for formal, specialised mechanisms to ensure quality standards and the provision of detailed information to stakeholders on outputs. Harvey (2006) defines quality assurance as the processes of 'establishing stakeholder confidence

that provision (input, process and outcomes) fulfils expectations, and measures up to threshold minimum requirements'. This definition draws attention to the importance of stakeholder confidence and the need for assessment mechanisms to ensure that quality thresholds and expectations of stakeholders are being met. Quality assurance also can be conceptualised as those 'policies, attitudes, actions and procedures necessary to ensure that quality is being maintained and enhanced' (Woodhouse 1999). Significantly, this definition draws attention to the importance of both the maintenance and enhancement of quality standards. In this volume, the previous chapter adopts an organisational effectiveness view of quality assurance, which is defined as a longitudinal rather than annual assessment that focuses on the multilevel structure of organisational effectiveness. It is based on both qualitative and quantitative data and seeks, primarily, to provide information for governments and higher education institutions. This definition emphasises primarily the assessment function of quality assurance and the role of the key stakeholders, especially governments and higher education institutions.

Accountability refers to 'rendering an account' about what an institution is doing in relation to goals that have been set, or legitimate expectations that others may have of one's services or processes, in terms that can be understood by those who have a need or right to understand 'the account'. For this reason, accountability is usually linked to public information and to judgements about the fitness, the soundness or level of satisfaction achieved (Middllehurst and Woodhouse 1995). Accountability raises important questions about who is accountable, for what, to whom and through what means? Romzek (2000) provides a comprehensive framework for analysing types of accountability relationships, identifying four basic types: hierarchical, legal, professional and political. All are important in higher education, although professional and political accountability are probably the most important reflecting situations 'where the individual or the agency has substantially more discretion to pursue relevant tasks than under legal or hierarchical types' (Romzek 2000: 25). Trow (1993) adds some useful points with regard to higher education. First, accountability should be seen as a constraint on arbitrary power, thereby discouraging fraud and manipulation and strengthening the legitimacy of institutions that are obligated to report to appropriate groups. Second, accountability sustains or enhances performance by forcing those involved to subject their operations to critical review. Third, accountability can be used as a regulatory device through the kind of reports and the criteria required by reporting institutions.

Both quality assurance and accountability have taken on much greater importance in modern higher education systems with the adoption of what may be referred to as the corporate model of governance under which universities are given enhanced autonomy to make their own decisions. This, in turn, requires governments to transfer substantial powers to universities and to steer institutions from a distance, using a variety of accountability mechanisms including: (a) quality assurance – where universities are held to account for the quality of their systems and outcomes; (b) performance-related funding – where funds are allocated on the basis of specified performance indicators; (c) market mechanisms – where universities are allowed to become more exposed to market forces, thus enabling users of the services

provided by them to be able to express their preferences directly through their purchasing patterns; (d) participation of external stakeholders on governing bodies; and (e) public disclosure of institutional performance requiring universities to make public details concerning, for example, the quality of teaching and of research, and the labour market outcomes of graduates (Santiago et al. 2008: 89). As Ministries divest responsibilities for governance, particularly with regard to the detailed institutional control and administration, they take on other responsibilities in terms of policy steering, and evaluation and accountability.

To some extent, different forms of quality assurance reflect different purposes. Indeed, Sachs (1994) has shown that, broadly speaking, quality assurance procedures can serve two major purposes: accountability and improvement. In this chapter, accountability is viewed as one purpose of quality assurance, although it is important to recognise that accountability often takes other forms, particularly related to legal and financial requirements. A recurrent theme in the literature relates to whether (and how) the purposes of accountability and quality improvement may be combined in a balanced strategy. On the one hand, some argue that accountability and improvement are incompatible as the openness essential for improvement will be absent if accountability is the purpose of the quality procedure (Woodhouse 1999). By contrast, others consider that accountability and improvement are closely linked and cannot be addressed separately, in which case the challenge for policy makers is to find effective ways of combining these two functions in the design of a quality assurance framework.

3.2 The Development of Modern Quality Assurance Systems

3.2.1 Drivers of Quality Assurance

Over the past two to three decades, modern quality assurance and accountability systems have been established in most countries. A number of factors have been important in driving these developments. First, modern quality assurance systems are common outcomes from the transition from elite higher education to mass higher education, leading to substantially increased student enrolments and expanded financial burdens on national governments. This has heightened the interest of governments in cost efficiencies and ensured that allocated funds are being well spent. In turn, much larger enrolments have often raised questions about whether or not academic standards are being maintained in times of rapid enrolment student expansion.

Second, as already noted, with the adoption of what may be referred to as the corporate model of governance, universities in many countries have been given enhanced autonomy to make their own decisions. But, in return, governments have set in place new quality assurance and accountability measures (Santiago et al. 2008: 89).

Third, the development of quality assurance can be seen as one consequence of the evolution of government in modern economies, with the implementation of

neo-liberal policy and its financial and administrative technologies in the New Public Sector Management (NPM), quickened by communicative convergence and global policy transfer (Marginson forthcoming). With a strong emphasis on achievement of efficiencies and the use of competition and markets, governments have understandably taken a growing interest in higher education, particularly public higher education or publicly supported higher education.

Fourth, with the growth of more knowledge-based economic activities highly dependent on information and communications technologies, associated with increased competition in international trade, there have been increased concerns about the need for large numbers of young people receiving higher-level qualifications relevant to labour market needs. This, in turn, has placed increased pressures on universities and colleges to produce graduates with appropriate knowledge and skills. Thus, quality assurance becomes an important mechanism to ensure that graduates have relevant, quality education to meet the needs of employers.

Fifth, quality assurance has become increasingly important with the internationalisation of higher education and increased student and professional labour mobility. Internationalisation not only places increased emphasis on greater supervision of the courses undertaken by international students but also has led to increased convergence of higher education systems. Professional bodies have been prompted to work to achieve sets of common national standards for professional recognition and so to facilitate the employment of graduates in foreign countries (Salmi 2008).

3.2.2 Establishment of Quality Assurance Systems in the 1980s and 1990s

The initial development of modern quality assurance systems took place simultaneously from three different but inter-related bases. First, in a number of European countries, under the threat of the establishment of government inspectorial functions, higher education institutions themselves established academic audits focusing on both institutions and courses. In Britain, in the late 1980s, for example, the Committee of Vice-Chancellors and Principals established an Academic Audit Unit to run institutional reviews since it feared that if it did not take action the Government would do so. This unit, in turn, was replaced by other non-government bodies with wider briefs, first the Higher Education Quality Council that carried out audits of both universities and polytechnics, and then in 1997 by the current Quality Assurance Agency (QAA). In the Netherlands in the 1980s, the Government agreed for the Association of Dutch Universities to establish and operate a system of separate national disciplinary reviews of both teaching and research.

This European model that soon spread to a number of countries had distinctive features. One was that the primary responsibility for quality rested with institutions themselves, while another was that audits focused primarily on the institutional-wide mechanisms of review for maintaining quality rather than on assessing quality directly. Prior to visits by external audit panels, institutions carried out self-studies

with reports being made available to panels. The purpose of external reviews was to ensure that standards were being maintained and enhanced, and that the institution had in place appropriate management structures for course approval and review (van Vught 1994).

The second base for the development of new quality assurance mechanisms was the adoption of American accreditation model by a number of countries that had traditionally close ties to the United States. Asian countries such as Korea and the Philippines were among the first to experiment with the accreditation model. In some cases, these new accreditation systems were organised by groups of institutions, whereas in other cases initiatives came from governments. Some systems had responsibility for particular groups of institutions, while in other cases the focus was the whole higher education system. Meanwhile, as international concerns with quality assurance developed, efforts were made to reform the American accreditation system by adding quality improvement to its key functions.

The American-style accreditation system has had well-established processes for making judgements about whether standards for a program or institution are being met. Following a request for accreditation, an institution completes an internal self-study that is followed by the visit of an external panel to ascertain whether or not minimum standards are being met. Such visits usually result in the granting of accreditation or conditional accreditation, although sometimes institutions are required to address particular problems in order to qualify for accreditation. Accredited institutions enjoy both recognition and accompanying status (Harman 1998).

Third, in the 1980s and 1990s, a number of countries began to experiment with new forms of evaluation, accountability and management, including student reviews of teaching, qualifications frameworks, regular program reviews, performance funding, links between peer review and funding on campuses and strategic planning and management. A particularly important role in these developments was played by American universities and American state governments (Rhoades and Sporn 2002). Many of these elements have been incorporated in the new more comprehensive national quality assurance frameworks.

Many of the ideas embedded in modern quality assurance systems are by no means totally new. For generations, universities and colleges in many countries have been concerned about academic standards and assessment, and have had in place various internal and external mechanisms to control requirements for student admissions and standards for the award of degrees and diplomas. In the United Kingdom, for example, universities have had long traditions of the use of external examiners in each discipline, while in the United States from the early twentieth century a strong system of institutional and program accreditation developed awarding official recognition to institutions and programs that met minimum standards. This American system was developed specifically to deal with problems of low academic standards among smaller late nineteenth century private colleges and it was a system established by groups of higher education institutions themselves, rather than by government.

What was new and important about the modern quality assurance movement, however, is that it introduced much more systematic and rigorous approaches to

assessing and ensuring the quality of courses and graduates. Such approaches included stronger emphases on regular internal reviews and reporting of results, more rigorous attempts to measure the academic standards and outputs, more serious efforts to gather the views of stakeholders and stronger commitments to making the results of assessments more widely available in the community. Stakeholders came to be defined to not only include students and their families but also employers, professional associations, alumni and the general public.

The new quality assurance arrangements that were put in place in the 1980s and 1990s showed considerable diversity and variety. In some countries, external reviews continued to be voluntary and were conducted by associations or agencies established by higher education institutions themselves, whereas in other countries reviews or audits became compulsory, and were carried out by government agencies. There were sharp differences too in the particular focus of reviews, especially whether they concentrated on institutional level quality assurance mechanisms or the quality of programs, and whether review processes focused primarily on teaching, or whether they also included areas such as research and community service (Harman 1998). There were also differences with respect to policy about who had access to review reports and whether or not funding allocations either directly or indirectly were linked to reviews.

3.2.3 Modern Quality Assurance Frameworks: United Kingdom and Australia

One important development in the 1990s was the transition in a number of countries from national quality assurance based on a single mechanism to multi-mechanism. National quality assurance frameworks made up a number of inter-related components at both national and institutional levels. Such frameworks generally specify and provide details of the various elements, as well as government and institutional responsibilities for collection and publication of nationally agreed data. Key characteristics of modern quality assurance frameworks are well illustrated by the cases of the United Kingdom and Australia. Details of the mechanisms included in each are summarised in the Table 3.1.

Table 3.1 Mechanisms included in British and Australian quality assurance frameworks

Mechanisms	British	Australian
Institutional audits	×	×
Qualifications framework	×	×
Accreditation of new providers		×
Codes of practice	×	
Subject benchmark statements	×	
Course experience questionnaire	×	×
Performance indicators		×
Assessment of research quality	×	

For two decades, the United Kingdom has been an important national leader in quality assurance with a strong commitment to its style of external institutional audits based largely on peer review. As already noted, under this system, universities and colleges have primary responsibility for quality assurance and must have their own well-developed internal quality assurance processes. The key national agency is the QAA, which is an independent body funded by subscriptions from universities and colleges and through contracts with the higher education funding bodies and carries out regular external quality assurance audits. In addition, largely in response to the recommendations of the Dearing Committee, the British quality assurance framework now includes codes of practice, subject benchmark statements, a qualifications framework and program specifications that are detailed statements specifying learning outcomes and how these may be achieved. In addition, since 1986, the Research Assessment Exercise (RAE) has provided periodic assessments of research on a disciplinary and institutional basis. While the RAE provides in depth assessments of research quality, it is not considered as part of the quality assurance framework although clearly it performs important quality assurance functions.

Since the late 1990s, Australia has had in place a multi-layered and comprehensive national quality system, with responsibility for setting and maintaining academic standards, and for internal quality assurance. External elements include:

> Regular audits of universities and State and Territory government agencies responsible for the accreditation of new providers by the Australian Universities Quality Agency which is an agency jointly owned by the national government and state governments
>
> Accreditation of new providers by State and Territory Government agencies, using agreed national protocols
>
> The Australian Qualifications Framework which specifies and defines recognised post-school and senior secondary school awards, and includes a national register of awards and recognised providers
>
> Supplementary mechanisms provided by the Commonwealth Government including outcomes data (Jones et al. 2001)

The Australian Universities Quality Agency (AUQA) conducts regular five-yearly audits of universities based on prior agreement between AUQA and the particular university on the major topics of focus. Audit reports do not give ratings or scores but, instead, provide reports about how well the institution is performing in assuring quality, commending some practices or aspects and making suggestions for improvement. Accreditation of new higher education providers is handled by state government agencies following agreed national protocols. State higher education agencies, in turn, are audited on a regular basis by AUQA. The Australian Qualifications Framework was developed in 1995 and classifies academic and professional qualifications, specifying the outcomes or competencies that should normally be achieved by graduates. During the 1990s, the Commonwealth Department of Education, Science and Training undertook a number of separate but related initiatives that, in the end, produced important supplementary elements of the quality assurance system, including a comprehensive system of performance indicators, advice on benchmarking and graduate attributes, and national data from the Course Experience Questionnaire, the Postgraduate Research Experience Questionnaire and tests of graduate skills. Particularly successful has been the Course Experience

Questionnaire that all graduates complete using a standard questionnaire in the year following their graduation. In addition, since the early 1990s, Australia has used simple quantitative indicators of total external research income, numbers of publications and number of higher degree completions to assess research quality, although it is doubtful whether this would count internationally as rigorous assessment (Harman 2004). However, the current national Government is in process of implementing a rigorous assessment system to be known as the Excellence for Research in Australia (ERA) initiative (Harman 2009).

3.3 Quality Assurance and Research

The research role of universities has had a somewhat uncertain place in modern quality assurance systems. While some quality assurance systems from the start, such as that in the Netherlands, have provided for separate national reviews of both teaching programs and research activities, others have paid relatively little attention to research and have concentrated attention primarily on teaching activities. Curiously, in the United Kingdom, the RAE that has been employed since 1986 to review the quality of university research has never been viewed as being part of the quality assurance system, although one of its explicit purposes has been to assess research quality. However, with the growing importance of research activities and outputs in both national innovation systems and various systems of university ranking, much more attention is now being given to research quality, whether or not such assessments are viewed as part of national quality assurance systems. Further, with substantial government investment in university research, new accountability pressures have emerged to ensure that investment not only leads to quality research but also that research outputs are valued by industry and the wider community.

Examples of three different systems of assessment of research quality are the cases of the Netherlands, the United Kingdom and New Zealand. The Netherlands system, which uses a protocol defined in cooperation with the universities, the Royal Netherlands Academy of Arts and Sciences, and the Netherlands Organisation for Scientific Research, has three main objectives: Improvement in the quality of research through an assessment carried out according to international standards; Improvement of research management and leadership; and Accountability to the higher levels of research organisations and funding agencies, government and society at large. Disciplines are reviewed sequentially with each having a different review committee. Evaluations are not used to allocate funds but to guide policy development (Geuna and Martin 2003). Institutes and university departments carry out self-evaluations containing both qualitative and quantitative information for the preceding 5 years. Visiting panels then make assessments based on four criteria: quality; productivity; vitality; and feasibility. Departments are evaluated on a five-point scale running from excellent to unsatisfactory.

For over 30 years, the British RAE has been by far the most thorough, in-depth and intrusive national assessment of research quality and outputs. From the start, it has been highly controversial, but there is a strong body of opinion within senior

levels of government and among vice-chancellors that, despite some limitations and adverse effects, the RAE has operated as a powerful assessment instrument focusing enhanced attention on research quality and outputs. The RAE was established by the former University Grants Committee as a response to the growing demands for public accountability and a more selective approach to block funding of research. Initially, it was planned to establish an assessment system for science-based disciplines only but, after consultation, it was agreed to also include the social sciences and humanities. RAE assessments were made in 1986, 1989, 1992, 1996, 2001 and 2008. Assessments used peer review methodology, with numerous disciplinary panels (some 70 by 2001) scrutinising publications as well as a basket of historical data on funding, research training and outputs submitted by higher education institutions about their research active units. Active researchers have been able to put forward for assessment up to four publications or other research outputs, with research quality until 2001 being was assessed on a seven-point scale (Geuna and Martin 2003). RAE assessments have been used by the research funding agencies to distribute substantial sums of block research funding. In 2009–2010, the Higher Education Funding Council of England (HEFCE) is allocating £1.572 billion to English universities on the basis of 2008 RAE results (Harman 2009: 159–162).

The New Zealand performance based research funding model introduced in 2004 is unique in that the individual researcher is the focal point of assessment. The aim of assessments is to improve the quality of research and public awareness of research, but in addition, assessments provide mechanisms for redistributing existing funding and allocating additional funding on the basis of the quality of researchers. In the original design, considerable notice was taken of the British RAE and the Australian Institutional Grants Scheme but it was decided to adopt a mixed model, with elements of both performance indicators and peer review. Under the formula adopted, a weighting of 60% was allocated for research quality based on peer review, 25% for research degree completions and the remaining 15% on external research income. Assessment of the performance of individual academic staff occurred within the context of 12 broad disciplinary groups. Quality is measured by the evaluation of individual "Evidence Portfolios" submitted by staff members who are assessed on a five-point scale. Following the first set of assessments, a Sector Reference Group was appointed to redesign the process (Tertiary Education Commission 2008: 8).

3.4 Recent New Directions in Quality Assurance

3.4.1 Drivers of the New Directions

Over the past 5 or 6 years, new policy currents appear to be substantially affecting quality assurance systems and higher education institutions. These new currents stem from an amalgam of factors including further shrinkage of public resources

for higher education, enhanced competition between providers, the rapid expansion of private providers and the impact of global institutional rankings. In addition, with many of the current national quality assurance systems having operated for a decade or more, governments are inclined to look critically at the performance of their quality assurance systems and question underlying assumptions that quality assurance should lie primarily with higher education providers rather than with governments. Particularly important results have been a convergence of accreditation and audit models, a more direct role for government in quality assurance, much closer attention to academic standards and student outcomes, greater transparency in terms of information provided to stakeholders, more rigorous assessment of research quality and impact, and stronger internal quality assurance mechanisms.

Here, we are particularly interested in the impact of global rankings on the quality assurance systems that developed over a two-decade period from about 1985. Clearly, global rankings are having substantial and sometimes unexpected effects. Particularly important has been the impact on institutional decision making. One recent study of the impact of rankings on institutions globally found that 63% of responding institutions were taking 'strategic organisational, managerial, or academic actions' in response to rankings, while only 8% indicated that they had taken no action (Hazelkorn 2007). Rankings also are influencing how institutions think about quality. According to a study by a Washington-based higher education policy institute, rankings 'appear to influence institutions' decisions in the quality assurance realm'.

> Research suggests that this impact is often positive; as a result of rankings, institutions may be motivated to improve the quality of their academic programs. Institutions report that lower than expected rankings are prompting curriculum review, new teaching and learning practices, and new academic programs (Institute for Higher Education Policy 2009: 3).

Even more important has been the impact on prospective student choice as reported by the 2009 OECD *Thematic Review of Tertiary Education* (Tremblay 2009) and by other studies that emphasise the impact of rankings especially on prospective international students (Thakur 2008). In addition, global rankings are affecting relationships between higher education institutions and governments. Research-intensive universities, for example, frequently base claims for additional resources on arguments about international competition and the need to maintain or improve global rankings. Research-intensive universities also use global rankings in their efforts to persuade governments to concentrate resources increasingly among relatively small numbers of leading universities.

Concern about global competition and rankings are prompting governments to tighten quality assurance and accountability requirements so as to ensure enhanced quality outputs and high institutional rankings. Governments also appear to want more control over quality assurance and quality assurance agencies. Significantly, the proposed major reforms to the Australian higher education quality assurance system have been justified in terms of new accountability needs and

to 'enhance Australia's position in international education' (Review of Australian Higher Education 2008: 115) and 'because Australia is losing ground against our competitors and unless we address (issues of quality and reputation) our prosperity will decline' (Lane 2009). In Britain, the recently retired Head of the QAA, Peter Williams in his Lord Upjohn Lecture in November 2009 accused the Government as trying to take over the QAA and use it for its own purposes. He saw this as a further example of enhanced state control over universities (Attwood 2009).

The ranking systems build on long traditions of informal status ranking of institutions that have been part of thinking about higher education. In addition, as Marginson notes,

> … some of the groundwork was laid by global communications, trade in education and NPM; and also the world-wide organization of science and research and the growing dominance of English as the one global language of research. Comparisons of national performance in education have been legitimated by the widespread use of cross-country comparisons in economic policy in which the nation state is seen to be rising or falling on the global plane. More specifically, university rankings have roots in normative methods of comparative education and development education, particularly the pervasive influence of comparisons of national policy performance by the OECD … and World Bank (Marginson forthcoming, 6).

3.4.2 Convergence of Audits and Accreditation and New Government Roles

The rapid new currents of change have led to the convergence of audit and accreditation methodologies and enhanced roles for national governments in quality assurance. According to Santiago et al. (2008), only a few OECD countries have continued with systems of external quality assurance based on academic institutional audits for all types of tertiary education institutions, while many countries that previously relied on accreditation processes have built in quality improvement aspects. Governments have been keen to strengthen quality assurance capabilities and, with the expansion of private higher education, to ensure that only those higher education institutions that meet threshold quality requirements are allowed to offer higher education courses. Associated with these changes has been an enhanced role for government in the quality assurance arrangements including more direct control by Ministers or Ministries of assessment or audit agencies. These changes reflect new pressures of accountability and concern about the quality of provisions and outputs.

Australia is an example of a country that is in the processes of making major changes to its quality assurance system along the lines described above. For almost a decade, as already noted, Australia has had a well-developed quality assurance system with quality audits of "self-accrediting institutions" being handled by AUQA and the accreditation of new providers being handled by the state higher education agencies following uniform protocols. Under proposals recommended

by the Bradley Committee (Review of Australian Higher Education 2008), a new national regulatory agency to be called the Tertiary Education Quality and Standards Agency will be established by the Commonwealth Government and will take over the responsibilities of both the AUQA and state higher education agencies. All new higher education institutions including public universities will need to be accredited in order to offer courses while all existing providers including universities will be periodically reaccredited. All providers will be required to undergo quality audits focused on institution's academic standards and the processes for setting, monitoring and maintaining them. This new agency will be a Commonwealth rather than joint State-Commonwealth body and in addition to audit and accreditation functions it will have responsibilities for comparative benchmarking and initiatives to improve student selection and retention, exit standards and graduate outcomes. Clearly, the intention is to create a quite powerful regulatory body, with a brief to be involved in many aspects of the operation of higher education institutions.

3.4.3 New Emphasis on Academic Standards and Student Outcomes

While the original emphasis with many quality assurance systems was on ensuring that institutions had strong management processes in place to guarantee quality, the new emphasis is increasingly to be directly concerned with academic standards and student outcomes. Recent developments in Britain and Australia illustrate this new direction.

In Britain, a substantial public debate about quality in higher education led in the summer of 2008 to an urgent parliamentary enquiry and later in 2008 to the appointment of a Teaching, Quality and the Student Experience Sub-Committee by HEFCE. The parliamentary enquiry took evidence from a wide range of stakeholders including Ministers, senior public servants, Vice-Chancellors, academic staff, students, the QAA and UK Universities with some substantial criticisms being voiced. In its report, the parliamentary enquiry was highly critical of evidence given by some university vice-chancellors and concluded that the QAA needed major reform. It observed:

> In our view, it was a matter of some regret – and a symptom of complacency – that it was only after pressure from outside the sector that the QAA used the cause for concern process to examine more generally institutions' capacity to assure the academic standards and quality of their higher education programmes and awards. We consider that the QAA needs to make up for lost time (House of Commons Innovation, Universities and Science and Skills Committee 2009: 216).

HEFCE's sub-committee was far less critical but it did admit that there are clear areas needing improvement including the external examiner system, information provided to students and the need for the QAA to change its focus to and take on

the role of explaining how and why the public should have confidence in English higher education (Higher Education Funding Council of England 2009b).

In Australia, as already noted, moves are under way towards a more outcomes-based approach that places much greater emphasis on academic standards and outcomes. The recent Australian Review of Higher Education recommended two particular possible strategies:

> Development of indicators and instruments to assess and compare learning outcomes directly; or
>
> Development of formal statements of academic standards by discipline for use in course design and approval processes as well as assessment and moderation processes across all institutions (Review of Australian Higher Education 2008, 134).

As already noted, the Commonwealth Government has accepted advice from the Bradley review that AUQA should be replaced by the Tertiary Education Quality and Standards Agency. This new agency will be a much more closely directed body and will have major responsibilities for academic standards and for the quality assurance framework.

3.4.4 Greater Transparency in Information Provided to Stakeholders

One of the most significant influences of global rankings has been its role in highlighting the inadequacies of meaningful information made available to stakeholders by quality assurance agencies and higher education institutions. Most quality assurance agencies publish reports of audit or accreditation visits, but these tend to be written in careful bureaucratic language and do relatively little to provide information to prospective students or employers. For example, the Australian Universities Quality Agency is most careful about offering major criticisms of institutions and structures its reports largely around the major headings of commendations, affirmations and recommendations. While higher education institutions often report brief summary data from surveys of student satisfaction and graduate destinations, they seldom provide access to the detailed studies on which the data were based.

The recent report of the HEFCE sub-committee on teaching quality and the student experience emphasised the importance of transparency and provision of helpful information to stakeholders. In his foreword to the sub-committee's report, the chair commented as follows:

> The need for information in a format that is easily understandable – and that corresponds what applicants, students, parents, employers and other stakeholders actually want to know – is paramount. This will oblige HEIs to provide information in a more standard, accessible format and will require significant changes to the QAA. It is essential, above all, that the QAA comes to regard itself as a public-facing organisation whose job is not only to assure quality, but also to explain how and why the public should have confidence in English Higher Education (HEFCE 2009b: 2).

3.4.5 *Stronger Emphasis on Research Assessment and Research Impact*

Another new direction is a much stronger emphasis being given to the assessment of research quality, with the introduction of new methods of assessment including the use of metrics and inclusion of assessments of research impact. The term "metrics" includes both bibliometrics and other quantitative data, such as total external research income or the number of higher degree completions. Bibliometrics are indicators of research performance based on citations in leading academic journals by other scholars, using the mainly the Thomson-Reuter Institute of Scientific Information (ISI) database. Use of citations is based on the assumption that more frequently cited items have greater significance and worth than those that remain uncited. Eugene Garfield, the founder of ISI, proposed that citation counts should seen as an index of impact, but more recently impact often has been interpreted as a measure of quality. While the main methods of assessment of research quality for the last two decades have been based mainly on peer review, there is a decided new move to combine the use of metrics with "light" peer review (Australian Research Council 2009; Hubbard 2008).

Both Australia and Britain are in the process of developing new systems of assessment of research quality based on metrics combined with peer review. However, in both cases it is proving difficult to develop arrangements that meet the support of academics across a range of different disciplines. Since 2006, HEFCE and Ministers have been engaged a consultative process to find a replacement for the RAE, which it is hoped will be less costly and less time consuming for institutions. The Government's intention was that the proposed Research Excellence Framework would depend largely on bibliometrics. However, it has proved difficult to achieve agreement about appropriate bibliometrics for non-science disciplines with the result that the Government made commitments that indicators of quality and impact will be used as appropriate to particularly disciplinary areas as determined by the advice of expert panels. Based on further extensive consultation, in September 2009, HEFCE issued a further consultation paper outlining a redesigned Research Excellence Framework in which a relatively small number of expert panels will assess research outputs, research impact and research environment (Higher Education Funding Council of England 2009a). Outputs will be assessed by peer review informed by citation data, while impact will be assessed using a case-study approach.

Similar developments have been under way in Australia since 2004 when the Howard Coalition Government commenced work to develop details proposals for a Research Quality Framework, which would have made extensive use of citation data and would assess impact using case-studies (Department of Education and Training 2007). Under the current Labor Government, the Research Quality Framework was abandoned and replaced by a proposed Excellence in Research for Australia Initiative that will assess research quality within the Australian higher education institutions using a combination of indicators and expert review by committees comprising experienced, internationally recognised experts. According to the most recent proposals, research will be assessed within eight broad disciplinary

areas on the basis of indicators which will include volume and activity, ranked outlets (i.e., standing of research publications), citation analysis (for those disciplines where at least half the total output of the discipline is indexed by citation suppliers), income, applied indicators, research esteem, and discipline specific indicators. Indicators will vary somewhat between the eight disciplinary areas. For example, research in the Physical, Chemical and Earth Sciences will be assessed on the basis of esteem indicators, applied indicators, citation analysis, research income, volume and activity analysis and ranked outlets (Australian Research Council 2009).

3.4.6 Strong Internal Quality Assurance

Another important recent development in many countries has been a much stronger emphasis being given to internal institutional quality assurance, including monitoring, review, assessment and reporting processes. While internal quality assurance processes are subject to review in accreditation reviews and academic audits, a number of countries now have legislative provisions requiring universities and colleges to have in place, their own processes of internal quality assurance. Such an approach is now followed by Belgium, the Czech Republic, Finland, Iceland, New Zealand and Poland (Santiago et al. 2008: 270). National agencies in many countries also provide advice on the design and implementation of local quality assurance frameworks.

Internal quality assurance has been strengthened by the introduction of a variety of approaches and methodologies. These include the use of institutional quality assurance frameworks with a number of inter-related elements, establishment of quality assurance centers or specialist groups located within teaching and learning centers, use of information based on student and employer surveys, performance indicator data drawn from student management systems, use of bibliometrics for assessing research quality, and benchmarking performance using formal benchmarking agreements with comparator institutions.

3.5 Conclusions

This chapter has outlined the development of modern quality assurance systems since about 1985 and considered particularly how, in the past 5 or 6 years, quality assurance systems have interacted with national and global ranking systems. Over a two-decade period from about 1985, new political pressures led many OECD countries to develop quality assurance systems based mainly on academic audit or accreditation methodologies. However, since about 2003 new political currents are leading to a substantial change in directions with the impact of ranking systems being of particular importance.

With the systems that developed largely in the 1980s and 1990s, a strong underlying assumption was that the primary responsibility for quality and standards

lay with institutions themselves. Generally, these new quality assurance systems were organised nationally by non-government bodies or by bodies with some distance away from direct Ministry control. Over time, a number of countries moved from quality assurance based on single element to multi-element quality assurance frameworks, which in the case of Britain included a qualifications framework, codes of practice, subject benchmarks and program specifications, while in Australia, the framework included a national protocol for the accreditation of new private providers by state government agencies, a national qualifications framework and supplementary mechanisms including surveys of student satisfaction and graduate destinations. In many cases, the assessment of research quality was not seen formally as part of quality assurance but such assessments became increasingly common and increasingly, they came to be used to allocate large sums of block research grants to institutions, or at least to review research directions.

Since about 2003, ranking systems combined with broader stakeholder concerns about quality and standards, and some degree of dissatisfaction with existing quality assurance systems, have produced new policy currents that are leading to substantial re-direction of quality assurance. Global rankings have become new forms of "super" quality assessment that have considerable attraction, with their use of simple numerical scores or league tables. They have confirmed the notion of a world university market in which higher education institutions are measured according to their relative standing on a global scale. Global rankings in particular are having a major impact on how institutions market themselves and their strategic directions. Global rankings influence student choice, especially for international students. In turn, these new influences combined with stakeholder concerns about academic standards are prompting governments to take a much tougher role in charting directions for quality assurance agencies and pushing agencies to address more directly issues about academic standards and student outcomes. Other important results are a convergence of academic audit and accreditation mechanisms, a new emphasis on transparency in the provision of information, new forms of assessing research quality and impact, and new emphases on internal quality assurance.

Since global ranking systems are relatively new, it is difficult to predict their longer-term development and their likely influence on quality assurance and accountability but clearly governments seem likely to take much stronger roles in directing quality assurance systems and focusing attention on academic standards. Quality assurance systems are likely to employ more sophisticated, quantitative methodologies, especially for the assessment of research quality and impact. On the other hand, challenged by global rankings and with increased stakeholder concerns about quality, universities are likely to lose their key roles as guardians of quality while there is likely to be increasing serious challenges to the role of quality assurance agencies. Because of the simplicity of their presentation of evidence, institutional ranking systems may well be accepted increasingly as providing more reliable ratings of university quality and reputation.

With an increasing number of firms and government agencies becoming actively involved in producing and publicising global and national rankings, increased

competition between providers of information seems inevitable. Already, Elsevier is challenging the domination of Thompson Reuters in providing citation data and significantly the Times Higher Education Supplement rankings in future will be generated by a partnership with Thompson Reuters, the current leading producer of citation information. National governments are likely become even closely allied with systems of national rankings, as has already happened in a number of European countries, while the European Commission is exploring the possibility of establishing its own system of global rankings. Competition among ranking systems could lead to confusion but most likely a small number of ranking systems will be able to maintain or establish their place in defining quality. This may well further erode the role and standing of quality assurance agencies.

For the future, considerable responsibility rests with the scholarly community not only to point to weaknesses in methodology and approaches of current rankings systems and warn of their dangers, but also to actively enter lively debate about the assessment of academic standards and the design and use of particular indicators. There is also need, as Tapper and Filappakou (2009) argue, for the scholars to explore what institutions see as the basis for their reputations and what changes they are prepared to undertake to sustain this.

References

Attwood, R. (2009). Engage with quality assurance or face state takeover, scholars told. *Times Higher Education*, 19–25 November.

Australian Government. (2008). Review of Australian higher education: Final report (2008). Canberra: Australian Government.

Australian Research Council. (2009). *Excellence in research for Australia (ERA) initiative*. Canberra: Australian Research Council.

Department of Education, Science and Training. (2007). *Research quality framework: Assessing the quality and impact of research in Australia: RQF submission specifications September 2007*. Canberra: Department of Education, Science and Training.

Geuna, A., & Martin, B. (2003). University research evaluation and funding: An international comparison. *Minerva, 41*, 277–304.

Harman, G. (1998). The management of quality assurance: An international review of international practice. *Higher Education Quarterly, 52*(4), 345–364.

Harman, G. (2004). Australia's system of quality assurance for higher education: Development, key elements and current operations. *Higher Education Forum* (Research Institute for Higher Education, Hiroshima University), *1*, 117–134.

Harman, G. (2009). National assessments of research excellence and impact: UK and Australian experience. In P. Clancy & D. Dill (Eds.), *The research mission of the university: Policy reforms and institutional response* (pp. 153–176). Rotterdam: Sense Publishers.

Harman, G., & Meek, V. L. (2000). Repositioning quality assurance and accreditation in Australian higher education. Evaluation and investigations program, higher education division. Canberra: Department of Education, Training and Youth Affairs.

Harvey, L. (2006). Understanding quality. In European Universities Association, Bologna (Ed.), *Handbook: making Bologna work*. Berlin: Raabe Academic publishers.

Hazelkorn, E. (2007). Impact and influence of league tables and ranking systems on higher education decision-making. *Higher Education Management and Policy, 19*(2), 87–110.

Higher Education Funding Council of England. (2009a). *Research excellence framework: Second consultation on the assessment and funding of research*. Bristol: Higher Education Funding Council of England.

Higher Education Funding Council of England. (2009b). *Report of the sub-committee for teaching quality and the student experience*, Bristol.

House of Commons Innovation, Universities, Science and Skills Committee. (2009). *Students and universities: Eleventh report of session 2008–09,* Vol. 1. London.

Hubbard, P. (2008). *HE consultation and our future: Research excellence framework: Consultation event 10, 11 and 17 January 2008*. Bristol: Higher Education Funding Council of England.

Institute for Higher Education Policy. (2009). *Issue brief: Impact of college rankings on institutional decision making: Four country case studies*. Washington, DC: Institute for Higher Education Policy.

Jones, E., Pearce, M., & Karmel, T. (2001). The Australian higher education quality assurance framework. In G Harman (Ed.), *Quality assurance in higher education: Proceedings of the international conference on quality assurance in higher education: Standards, mechanisms and mutual recognition*, Bangkok, Thailand, 8–10 November 2000. Bangkok: Ministry of University Affairs of Thailand and UNESCO Principal Regional Office for Asia and the Pacific, 116–130.

Lane, B. (2009). Reform rationale. *The Australian*, 18 November.

Marginson, S., & van der Wende, M. (2006). To rank for be ranked: The impact of global rankings in higher education. *Journal of Studies in International Education, 11*(3/4), 306–329.

Middllehurst, R., & Woodhouse, D. (1995). Coherent systems for external quality assurance. *Quality in Higher Education, 1*(3), 257–268.

Rhoades, G., & Sporn, B. (2002). Quality assurance in Europe and the U.S.: Professional and political economic framing of higher education policy. *Higher Education, 43*(3), 355–390.

Romzek, B. (2000). Dynamics of public sector accountability in an era of reform. *International Review of Administrative Sciences, 66*(1), 21–44.

Sachs, J. (1994). Strange yet compatible bedfellows: Quality assurance and quality improvement. *Australian Universities Review, 37*, 1.

Salmi, J. (2008). *The challenge of establishing world-class universities*. Washington, DC: The World Bank.

Santiago, P., Tremblay, K., Basri, E., & Arnal, E. (2008). *Tertiary education for the knowledge society, Vol. 1 Special features: Governance, funding quality*. Paris: OECD.

Stella, A., & Woodhouse, D. (2006). *Rankings of higher education institutions*. Melbourne: Australian Universities Quality Agency.

Tapper, T., & Filappakou, O. (2009). The world class league tables and the sustaining reputation in higher education. *Journal of Higher Education Policy and Management, 31*(1), 55–66.

Tertiary Education Commission. (2008). *Performance based research fund: 2012 quality evaluation*. Consultation paper 1. Wellington: Tertiary Education Commission.

Thakur, M. (2008). The impact of ranking systems on higher education and its stakeholders. *Journal of Institutional Research, 13*(1), 83–96.

Tremblay. (2009). OECD: Diverse processes, common objectives. *University World News*, 4 May.

Trow, M. (1993). Managerialism and the academic profession: The case of England. *Studies of Higher Education and Research, 4*, 1–23.

Usher, A., & Salbvino, M. (2007). A global survey of university ranking and league tables. *Higher Education in Europe, 32*(1), 5–15.

Van Vught, F. (1994). Western Europe and North America. In A. Craft (Ed.), *International developments in assuring quality in higher education*. London: Falmer.

Woodhouse, D. (1999). Quality and quality assurance. In *Quality and internationalisation in Higher Education.* Paris: OECD.

Chapter 4
Social Contexts and Systemic Consequence of University Rankings: A Meta-Analysis of the Ranking Literature

Ulrich Teichler

4.1 The Late Arrival of the Ranking Boom in Europe

In the blurb of a recent book by two European higher education researchers on university rankings (Kehm and Stensaker 2009), we can read the following: "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 to the rest of the world, and that we also have seen 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 text reflects the fact that the existing differences in "quality," "reputation," "performance," etc. between universities within an individual country have not been viewed as a very important issue in many European countries until recently:

- On the one hand, the vertical differences between the universities within various European countries were perceived as so small that they were not highly relevant for those intending to enroll (e.g., in Germany and the Netherlands), for policy makers supervising and funding higher education, or for employers recruiting graduates.
- On the other hand, wherever vertical differences were considered, for example, in some European countries (such as the United Kingdom or France, especially with regard to the Grandes Écoles) or as regards certain purposes (e.g., research cooperation between university and industry or recruitment for the academic profession) in countries with a flat reputation hierarchy, informal "knowledge" about such differences was viewed as sufficient to make appropriate decisions.

U. Teichler (✉)
International Centre for Higher Education Research Kassel (INCHER-Kassel)
University of Kassel, Kassel, Germany
e-mail: teichler@incher.uni-kassel.de

J.C. Shin et al. (eds.), *University Rankings*, The Changing Academy – The Changing Academic Profession in International Comparative Perspective 3,
DOI 10.1007/978-94-007-1116-7_4,

The growing conviction in Europe from the 1960s onward that diversification of higher education was needed to cope with the increasing student enrolment rates and rising costs of research has not led to greater emphasis on quality differences between individual universities. Rather, a formal diversification of higher education was realized whereby some sectors, in many cases specific types of higher education institutions, had no or at most a marginal research function and teaching in these sectors had a more applied emphasis (e.g., for some period, the polytechnics in the UK and subsequently the Fachhochschulen in Germany and later also in Austria and Switzerland, the IUT in France, the HBO in the Netherlands, the AMK in Finland, etc.).

In analyzing the development of systems of higher education, the author of this chapter argued that "vertical" differences of "quality" and "reputation" between individual universities and departments began to play a visible role in Europe in the 1980s, and attention to this phenomenon constantly increased thereafter from about the 1990s (Teichler 2007a). Only when world-wide rankings became a popular issue of debate in the first years of the twenty-first century, did rankings really become a "power" all over Europe.

This shows how the knowledge on higher education worldwide, even among experts, is shaped by changing "problem awareness" or "fashions" in the higher education policy. Most higher education experts in Europe are not aware of the fact that university rankings traditionally play a more important role in East Asia than in the USA and that this was already true before the age of "massification of higher education."

The author of this chapter wrote his doctoral dissertation on higher education in Japan between 1970 and 1974. The initial aim was to analyze how the characteristics of study programs and graduate employment and work change as a consequence of the rapid increase in student enrolment on the way to a "highly educated society." It was shown that, in Japan, the concept of a highly educated society (kôgakureki shakai) is invariably linked to that of "over-competition" of a highly vertically stratified higher education system where rankings – in Japan, notably the selectivity in admission to universities – play an important role as an information and steering tool. In the dissertation published as "Das Dilemma der modernen Bildungsgesellschaft: Japans Hochschulen unter den Zwängen der Statuszuteilung" (The dilemma of the modern education society: Japan's higher education institutions under the forces of status distribution) (Teichler 1976), the author pointed out that the trend toward an educational meritocracy, i.e., an increasing belief that education is open to everybody and crucial in determining future societal, professional and life chances, leads to a fierce competition for small distinctions of education attainment because modern societies keep inequities of status (it was called – as an ironic response to the concepts of manpower demands – "demand for social inequality") and more and more people do all they can to seek the highest possible success. Obviously, the effects of the combination of "over-competition" prior to entry to higher education and low achievement of the university students as a consequence of exhaustion suggest that a boundless educational meritocracy faces the dilemma that the potential virtues of expanded education are undermined. One might even conclude that only a "moderate meritocracy" can work.

In comparing the debates on rankings in Japan and the USA in the 1970s, the author concluded that the social relevance of rankings in Japan was much greater than in the USA; smaller proportion of future university students in the USA than in Japan was eager to absorb such information. This seemed to be related to the fact that opportunities to reach educational success were seen as more open in Japan and that professional success and status were seen as more strongly determined in Japan than in the USA, where the dream of career success "from the dish washer to the millionaire" left hope for life-long corrections of a mediocre career start (Teichler 1991).

In re-visiting university rankings now, a few decades later, after this issue became fashionable worldwide and was put on the agenda by what one might call a "ranking movement," the author continued to be interested in the effects of rankings beyond the heroic intentions claimed by their advocates (e.g., transparency, healthy competition, and overall quality and efficiency improvement). These effects should not be called "unintended consequences," because we cannot distinguish clearly between "intended" and "unintended" consequences. The activities to establish rankings are deeply intertwined with the belief of the virtue of steeply stratified higher education systems. This belief is characterized by a world view that small differences matter enormously and that some of the effects others migt view as undesirable are viewed as desirable by the advocates of ranking. For example, the Norwegian social scientist Galtung (1971) pointed out that, although there was a widespread complaint in Japan about the undesirable side-effects of "over-competition," many Japanese considered it desirable that ranking and competition reinforce a "gambare" culture – what he called "effortism": It is desirable that everybody tries as hard as possible to achieve what she or he can.

There is another misunderstanding in the text quoted initially. It was said: "University rankings are a relatively new phenomenon in higher education." But even in a country like Germany where we noted a very flat hierarchy, rankings are by no means new. A substantial number of ranking studies were undertaken in the 1970s and the 1980s (cf. the overview by Teichler 1986). Some of these were discussed for a few days in the press, but they have not stirred up major public debates for a long time. Only since the 1990s, have they been "re-invented" and become very popular. The rankings are not new in Europe, only their socio-political "effectiveness" or their power of distortion is new.

4.2 The Aim of This Analysis

The following meta-analysis of "rankings" addresses the effects of rankings beyond their often stated objectives. It focuses on the question: How are higher education systems transformed as a consequence of rankings?

In theory, one could divide this into two separate questions: What transformations are more or less deliberately pursued as the consequences of the objectives, the hidden agendas, and the methods actually employed? What is the unforeseen impact?

As has already been pointed out, a clear distinction between these two questions is not possible. It is not clear what we could consider as intended or as hidden agendas of the rankings studies. Nor is it clear, as discussed above, whether certain consequences are not unforeseen impacts, but rather desired and intended consequences on the part of the advocates of rankings.

This analysis was undertaken with the support of some master students in a course on "Higher Education Systems" in the International Master Program "Higher Education" offered by the Department of Social Sciences and the International Centre for Higher Education Research at the University of Kassel (Germany). Their involvement not only helped to analyze more than 50 publications discussing the concepts, methods and results of ranking studies, but also ensured that the views and experiences as regards rankings from different cultures could be reflected in this account. I am indebted to Elena Aleinikova-Schimmelpfennig (Belorussia), Thi Thanh Hoa Bui (Vietnam), Kwan Heung Lam (Hong Kong), Robert Keith Owino Odera (Kenya), Thi Lan Phuong Pham (Vietnam), and Xiang Yi Tao (China).

4.3 The Dominant Talk and Undercurrents as Regards Rankings

We can embark on an intellectual debate about rankings: What are the underlying rationales, methods, findings and impact on the higher education system and beyond? But this would not enable us to understand the current tones of passion in the ranking debate. Rankings have an enormous "sub-intellectual" or "extra-intellectual" power in the public discourse.

The first "sub-intellectual" element could be: Rankings are "sexy." They arouse feelings of doing something which leads to satisfaction. Although these feelings are not fully approved in the high-culture of our official communication, there are elements of a peep-show; they stir up feelings of shame and desires for boasting, and the like.

The second sub-intellectual element is the openness of the arena of information producers to unlimited low quality. Testing learners' abilities requires a minimum competence and a certain degree of acceptance between producers of information and those who supervise and run educational institutions. As a consequence, a discourse about standards of testing can work. University rankings, however, can be produced without any communication with those supervising and running the system, and the simplicity of concepts and methods does not seem to be an impediment for popularity.

The third extra-intellectual element is not only indicative for the ranking discourse, but also for some other discourses in higher education. The producers and advocates of the issue at stake invest so much time and energy in ruling the debate that discourse is dominated by the lobby and the critical voices have little chance of being heard.

In various other areas of higher education policy, we also note relatively open arenas of policy discourse where groups of persons made up of actors of a certain policy and activity, in this case producers of ranking studies, advocates and "experts" sympathizing with the respective policies and activities, meet. For example, we note similar circles as regards "quality assurance," funding, and student mobility. This is certainly helpful to ensure a forum of constructive critique; critical voices are included insofar as they share the basic philosophies, thus contributing to the generation of ideas for the enhancement of respective activities. But such arenas and networks built up in this framework can serve to create a lobby for the respective activities, and can claim in the public policy debates that there is a conventional wisdom about the strengths of these activities and the challenges for improvement. In the case of university rankings, UNESCO-CEPES (Centre Européen pour l'Enseignement Supérieur) played an active role in stimulating such a discourse among actors, advocates, and sympathizing experts and in establishing the "International Ranking Expert Group" (Sadlak and Liu 2007).

The fourth extra-intellectual element of the ranking debates is the strong emotions that come into play on the part of the key players of the discourse. We note a "movement" in favor of rankings by the key producers and advocates as well as a congregation of "concerned scholars" in the critique of rankings. Again, this is not unique to rankings; many recent higher education reforms are advocated through major campaigns. Though values can be helpful to increase the depth and clarity of discourse, the emotions stirred up in the ranking discourse do not seem to serve intellectual clarity.

At present, we find three major types of literature on university rankings:

- *Publications by actors and sympathizers of the "ranking movement."* They share certain underlying values as regards higher education, a belief in the benefit of producing rankings for higher education and its environment, and see critical voices concerning the methodology of rankings as challenges for further investments in rankings and quality improvement of rankings.
- *Publications by fundamental critics and skeptics of rankings.* As a rule, they address the underlying concepts and the effects of rankings without any detailed discussion of the methods, data, and research findings of the effects.
- *Higher education researchers*. This is a small group who examines the key issues of the discourse. They analyze the concepts, methods, data, and effects without any a priori positive affiliation or negative dissociation from the "ranking movement." Dill (2009), Hazelkorn (2008), Marginson (2008), and Usher and Savino (2006) could be quoted as examples.

In the available literature, three major themes are very frequently discussed: (a) the concepts underlying ranking studies: on the one hand, the declared objectives and on the other, the hidden rationales and "ideologies"; (b) the methods of analysis: the criteria, the measurements, the data validity, the quality of the data, the presentation of the findings, etc.; and (c) the effects of ranking studies in transforming the higher education system.

4.4 Rationales and Hidden Agendas

4.4.1 Emphasis on Widely Shared Goals

In studies which advocate rankings, more or less uncontroversial aims are put forward, i.e., aims which should be accepted by anybody except those who do not care about quality, achievement, and progress:

- Rankings contribute to transparency which is called for because one assumes that many stakeholders yearn for good information and are in need of good information systems in order to be rational actors
- Rankings locate institutions or departments on "the map" of higher education in a fair way
- Rankings are an information system serving the idea that the best achievers will be rewarded: ranking is an integral element of a meritocratic system
- Rankings reinforce virtuous, healthy competition; the information on rankings has an overall stimulating effect of increasing efforts to improve

We know that these claims deserve to be challenged. Yet, it seems appropriate to state that the advocates of rankings like to put forward the benefits of rankings which, in principle, should be seen as desirable by everybody.

4.4.2 Underlying Rationales and Hidden Agendas

In ranking studies, "intentions," "rationales," "hidden agendas," and "ideologies" are in play but these are neither explained nor reflected by the advocates and producers of rankings. In some cases, they emphasize additional objectives, but most of the items in the following list are mainly put forward by critics of rankings as underlying hidden agendas or ideologies of the advocates and producers:

- An increase in fierce, rivalry-dominated competition.
- More strongly competitive, entrepreneurial and extrinsically motivated scholars.
- The growing role of worldwide competition.
- The relevance of "flagships" and "skyscrapers" in an increasing global competition among nations.
- An emphasis on vertical differences. Vertical differences are important, and horizontal differences are substantially less important ("compelling popularity of vertical diversity," see Marginson 2008; Marginson and Van der Wende 2007).
- A steeply stratified system of higher education is desirable.
- There are common world-wide criteria for a good university.
- There is a close link between the criteria of academic quality and those put forward in the name of the "knowledge economy."

- The quality of teaching and research of individual scholars or within individual units of a higher education institution is strongly determined by the institutional environment.
- A strong concentration within a few universities of high quality helps to increase the overall quality of the higher education system; it is not a "zero-sum-game" (see the critique by Deem, Lucas, and Mok 2009).
- Rankings favor inter-institutional diversity and run against intra-institutional diversity (Hughes 2008).

Additionally, we observe many other issues in this context. A few examples show that the above list is not exhaustive.

- Research is clearly viewed as the dominant driver and the major indicator of the quality of higher education (cf. the critique by Deem et al. 2009; Marginson and Van der Wende 2007).
- The desire to gather new knowledge, curiosity, etc. is deeply intertwined with the "desire to be first" (Sadlak 2007; Zhao 2007).

4.5 Self-critical Dimensions of the Ranking Advocacy

The advocates of rankings do not claim that rankings are perfect and that all their effects are desirable. Rather, they often quote selected aspects of the widespread methodological critique:

- Rankings are strongly shaped by the availability of data.
- There is no desirable balance between input, process, and output indicators.
- There are typical weaknesses of "subjective" or "objective" measures.
- The aggregate institutional performance does not sufficiently take into account differences by discipline and often overstates the role of science and engineering.
- There are problems of reliability of measures over time, e.g., ups and downs of positions of certain individual universities.
- There are problems of reliability across ranking studies, e.g., the unlikely high position of some universities in one ranking study and the unlikely low position in another.

However, the advocates of rankings tend to claim that these methodological weaknesses are not a real threat to the presumed benefits of rankings:

- Either the weaknesses are seen as a "necessary evil" of condensing information into indicators.
- Or the methodological weaknesses are interpreted as a challenge for future improvement. More money and energy should be put into ranking studies to improve their quality.

We often see claims by the advocates of rankings that more or less all the critiques are of a methodological nature. Therefore, the critique is also interpreted as being indirectly a call for the future improvement of rankings.

We also note that advocates of rankings consider critique of a systematic nature as justified. For example:

- The use of the information on rankings is uneven, e.g., notably by elite students.
- Networks of elite universities could lead to oligopolistic advantages.
- A healthy competition among universities for advancement could be impeded by varying degrees of institutional autonomy of the universities (Salmi and Sayoran 2007).

The advocates of rankings usually see these types of phenomena as current imperfections of market mechanisms or as imperfections of rational-actor behavior which would fade away if the conditions for a real market and for rational action were improved with the help of a good information system. Therefore, the weaknesses are not viewed as a fundamental challenge of the world-view of the protagonists of rankings.

Finally, we often hear the argument that everything should be done to improve rankings because they "are here to stay." Hence, any fundamental critique is futile and efforts of improvements are the only possible rational response to observed weaknesses. This argument is by no means convincing, because we are not inclined to accept negative aspects in other areas of life, e.g., drug consumption, mafia, corruption, etc., just because they are "here to stay."

4.6 Critical Voices: The Nine Major Endemic Weaknesses of Rankings

The advocates claim that the critics mainly emphasize the methodological weaknesses of rankings, which, from their point of view, are either unavoidable limitations of indicators or could be redressed by future improvements. The critics, in contrast, argue that the limitations of the rankings – no matter whether conceptually intended, caused by subconscious ideologies, unintended consequences, caused by methodological differences, or whatever other reason – lead endemically to undesirable effects. What are the major critiques?

4.6.1 The Vicious Circle of Increasing Distortion

The most frequently employed argument about the endemic malfunction of rankings is the following: Whatever the imperfections, the actors in the system will try to succeed according to the indicators. This:

- takes the means of measurement seriously as if they were the ends of quality in higher education (Bowden 2000; Dill and Soo 2005; Hazelkorn 2008)

- magnifies the distortions because actors try to be successful according to the distorted measures. As a consequence, the measures indicate even more distortions and the universities try to be more successful according to the more distorted measures, etc.

4.6.2 *Endemic Weaknesses of Data and Indicators*

There is a second widespread argument. The so-called methodological weaknesses do not coincide. These could be alleviated or eradicated by more investment in methodology and data collection, but they "are here to stay":

- Indicator-based assessment is bound to be under-complex (Altbach 2006; Marginson 2008; Hazelkorn 2008). So, if universities take rankings as challenges for improvement, their performance is bound to lose the complexity which higher education has to serve in order to contribute to the knowledge system and society.
- The high socio-political power of indicators invites cheating in the production of data.
- The rankings are more likely to be driven by the availability of data than by any concept of desirable indicators and data. This holds true most dramatically for world-wide rankings because it is almost impossible to influence world-wide data collection (in contrast to national data collection) by strategic decisions regarding data improvement – especially, if the data collection cannot be done without the help of the various countries.
- Collecting information to support rankings, produce alternative data to rankings and to respond institutionally to ranking is time-consuming and a burden. Rankings increase the administrative load. As this cannot increase forever, there will always be counter-pressures for simplification.

4.6.3 *The Lack of Agreement on "Quality"*

There is no minimum consensus about quality (Usher and Savino 2006):

- The extremely varied concepts of the existing ranking studies are not just the consequence of the fact that everybody can easily produce a bad ranking. We even note extremely divergent concepts and hidden agendas of quality in conceptually and methodologically ambitious rankings.
- It seems to be extremely difficult for rankings to respond to the typical discourses on the homogeneity or diversity of the concepts of quality, e.g., "fitness for purpose" versus "fitness of purpose," "quality according to internal standards of disciplines" versus "relevance," quality in disciplinary terms versus quality in interdisciplinary terms, etc.

- Most ranking studies are "monotheistic" in the belief that there is a single "god" of quality. There are hardly any ranking studies that deal with diverse concepts of "quality" or even "qualities."
- In the framework of evaluations of study programs or of assessments of individual scholars or research groups, or even of institutional accreditation, both the endemic weaknesses of indicators and the lack of consensus on quality are compensated by peer review processes (no matter whether they resolve the problems or only cover them up). In rankings of departments and universities, however, such a counterbalancing mechanism does not exist.

4.6.4 Imperialism Through Ranking

Rankings systematically favor universities of certain countries and thereby propagate them as the role models for those in other countries. This often happens through the reproduction of national ideologies in rankings (cf. Deem et al. 2009). Some examples are:

- The preoccupation of Chinese rankings to declare success in research as more or less automatically determining success in teaching and learning.
- The preoccupation of British rankings to consider inward mobility as an indicator of quality, neglecting the contribution of outward mobility to quality.
- The English-language bias of many international rankings discriminates against high quality work in countries that do not follow the lingua franca-dictatorship; ignorance of academic work in non-English languages (Salmi and Sayoran 2007).
- Countries with different configurations of national higher education systems are pushed to imitate the configuration of countries which were the initial breeding ground of rankings (i.e., those with a steeply stratified higher education system, cf. below, see Teichler 2007b). We note that some countries try to enhance their national prestige in a similar way by building skyscrapers (Zhao 2007).
- Rankings discriminate against universities in developing countries.

4.6.5 The Systemic Biases of Rankings

Rankings are unfair to this existing diversity of higher education. The following biases are mentioned most frequently:

- Rankings miss (or negatively assess) higher education institutions with other functions than those of research-oriented universities.
- Rankings undermine horizontal diversity and there is a disregard for diverse missions (Proulx 2007) and "fitness for purpose."
- Rankings disregard/disadvantage small institutions.
- Rankings discriminate against humanities and social sciences.

- Rankings reinforce dominant paradigms, thereby controlling the choice of theories, methods, and themes (Marginson 2008).
- Rankings do not sufficiently strike a balance between teaching and research, but often only infer that good research produces good teaching, that input in teaching and research leads to good processes, and that this in turn leads to good outcomes. Available research has often challenged these assumptions (see Dill and Soo 2005) and experience has shown that the "most successful institutions may not have the best practice" (Zhao 2007: 326–327).

4.6.6 Preoccupation with Aggregates

Rankings aim to measure the average quality of a higher education institution or of a department. However, we note scholars or teams of scholars within these aggregates who differ substantially in quality from the average. Often differences between high- and low-achievers within a single institution or department are far more striking than those of the average achievement of scholars between institutions or between departments. Measuring the average quality of scholars of the institutions or departments that are being compared seems to be based on the assumption that the quality of the academic work of individuals is strongly influenced by the quality of their local neighbors. This is advocated at a time when scholars can communicate world-wide with outstanding partners in other institutions more easily than ever before because of such things as cheap travel, the internet, etc.

Rankings discriminate against and eventually help to penalize outstanding scholars or units whose peers at their institution are not as outstanding as they are (Teichler 2007b).

4.6.7 Praise of and Push Toward Concentration of Resources and Quality

Rankings seem to be based on the beliefs that:

- A high concentration of resources in a few institutions is beneficial.
- Homogeneous teaching and research milieus (e.g., the entry selection of highly similar students) are desirable.

Thus, rankings encourage investment in a few outstanding universities (Marginson and Van der Wende 2007; Salmi and Sayoran 2007; Hazelkorn 2008, however, notes exceptions).

As a consequence, rankings encourage outstanding scholars to move to institutions with many outstanding scholars or units. Finally, rankings are based on the assumption that inter-institutional diversity is desirable and intra-institutional diversity is detrimental.

4.6.8 Reinforcement or Push Toward Steeply Stratified Systems

Rankings not only draw the attention of actors, observers, and "customers" to vertical differences but also discourage and possibly undermine horizontal diversity. They also favor steeply stratified institutional patterns. This is done through:

- Encouraging resources at a few outstanding institutions
- Making newcomers synonymous with losers
- Reinforcing mechanisms whereby status breeds status; there is the "Matthew effect" in resource allocation, and "reputational recycling"
- Undermining the pride of institutions which are not top institutions

Hence, it is not surprising to note that the institutions that are the winners of an increasing concentration of resources improve in quality. But there is hardly any evidence that the "quality" of the total system of higher education in a given country is enhanced if resources are highly concentrated at a few institutions and if the overall system becomes more highly stratified than in the past. We only know that the strengths of less stratified higher education systems (e.g., more mobility within the system, more contributions to regional development, easier access to high quality education for disadvantaged groups, etc.) are in danger of getting lost.

4.6.9 Rankings Undermine Meritocracy

Many experts are convinced that rankings undermine meritocracy (e.g., see Clarke 2007) and thus eventually endanger the quality gains often hoped for. The following anti-meritocratic effects are often quoted:

- Rankings reinforce the reputation of old universities and thus reward past rather than present achievements (Altbach 2006; Salmi and Sayoran 2007; Marginson 2008; Teichler 2007b).
- Rankings encourage the non-meritocratic selection of students (Clarke 2007); one tries to "purchase" the most likely winners according to views of predictive validity by socio-biographical background rather than by visible study potentials.
- Similarly, opportunities increase for students to buy status through particularistic admissions.
- The expected beneficial effect of being admitted to a high-ranking university could lead students to become complacent once they are enrolled.
- Rankings reinforce high entry standards and high exit standards, but not high "value added."
- As already pointed out, rankings tend to discriminate against outstanding scholars who are not surrounded by other outstanding scholars.
- Encouragement of strategies of higher education institutions to keep or to improve their rank does not always lead to quality gains: "the most affluent, elite

universities engage in a kind of functional inefficiency to maintain their status" (Bowden 2000: 42).
- There are increasing examples of data cheating and of getting unfair favors in the competition for resources; over-heated competition is likely to lead to deviant competitive behavior.

4.7 Conclusions

The production of rankings is a completely open arena because everybody can produce and disseminate primitive and highly ideological information. This is in contrast to educational testing where high costs and limited access to information ensure certain minimum standards.

The popularity of rankings is not positively related to their "quality" and normative "acceptability." There is even a "law" of a negative correlation, namely, the lower the quality and the more biased the normative basis, the higher the popularity of a ranking study is likely to be.

An analysis of the stated objectives of rankings is not enlightening. Statements of objectives are largely part of a "justification game" in a hotly debated and controversial setting of discourse. Moreover, it is impossible to establish how seriously the stated goals stated are pursued, what the disguised goals are, and how "ideologies" unconsciously come into play.

Consequently, we cannot distinguish clearly between intended and unintended consequences. Moreover, views vary as regards what consequences are desirable or undesirable.

The weaknesses admitted by advocates of rankings and those put forward by critics overlap. However, they are interpreted differently. What the advocates of rankings consider as acceptable weaknesses or as avoidable through methodological improvement and increase of rational-actor behavior, the critics often consider as endemic distortions.

In cumulating the biases of ranking, we can argue that the rankings favor not only English-speaking, research-intensive institutions with strengths in natural sciences, but also large, older institutions in countries with long ranking traditions, in countries with steep hierarchies and with little intra-institutional diversity.

There are nine major arguments as regards the endemic weaknesses of rankings: (a) the vicious circle of increasing distortion, (b) endemic data weaknesses, (c) lack of agreement on "quality," (d) "imperialism" through rankings, (e) systemic biases of rankings, (f) preoccupation with aggregates, (g) push toward concentration of resources, (h) push toward steep stratification, and (i) anti-meritocratic effects.

There are no indications that rankings will fade. Efforts to improve the quality of rankings here and there cannot prevent the constant production of less methodological "ambitious" rankings, the popularity of "stupid" rankings and the vested interests in producing "distorted information." One can only hope that the critique of rankings prevents an all-pervasive distortion terror of the rankings.

Finally, there are open issues which call for answers: How strong are the effects which are claimed by the advocates and the critics of rankings?; Is the future knowledge society best reinforced by a steep stratification of knowledge or by a broad, "flat" distribution of knowledge?; Are we moving toward greater meritocracy, do we settle and have to settle with a "moderate" meritocracy, or is meritocracy losing out vis-à-vis new increasing powers of information manipulation (are we moving toward a "de-information society?") and new modes of achieving and securing privileges?

References

Altbach, P.G. (2006). The dilemmas of rankings. *International Higher Education*, No. 42, 1–2.

Bowden, R. (2000). Fantasy higher education: University and college league tables. *Quality in Higher Education, 6*(1), 41–60.

Clarke, M. (2007). The impact of higher education rankings on student access, choice, and opportunity. *Higher Education in Europe, 32*(1), 59–70.

Deem, R., Lucas, L., & Mok, K. H. (2009). The world-class university in Europe and East Asia: Dynamics and consequences of global higher education reform. In B. M. Kehm & B. Stensaker (Eds.), *University rankings, diversity and the landscape of higher education* (pp. 117–134). Rotterdam: Sense Publishers.

Dill, D. D. (2009). Convergence and diversity: The role and influence of University rankings. In B. M. Kehm & B. Stensaker (Eds.), *University rankings, diversity and the landscape of higher education* (pp. 97–116). Rotterdam: Sense Publishers.

Dill, D., & Soo, M. (2005). Academic quality, league tables, and public policy: A cross-national analysis of university rankings. *Higher Education, 49*(4), 495–533.

Hazelkorn, E. (2008). Learning to live with league tables and ranking: The experience of institutional leaders. *Higher Education Policy, 21*(2), 193–216.

Hughes, R. (2008). *International collaboration and the effect of rankings*. Nottingham: University of Nottingham.

Galtung, J. (1971). Social structure, education structure and life long education: The case of Japan. In *Reviews of national policies for education: Japan* (pp. 131–152). Paris: OECD.

Kehm, B. M., & Stensaker, B. (Eds.). (2009). *University rankings, diversity and the landscape of higher education*. Rotterdam: Sense Publishers.

Marginson, S., & Van der Wende, M. (2007). To rank or to be ranked: The impact of global rankings in higher education. *Journal of Studies in International Education, 11, 3–4*, 306–329.

Marginson, S. (2008). *The new world order in higher education: Research rankings, outcome measures and institutional classifications*. Victoria: University of Melbourne, Centre for the Study of Higher Education.

Proulx, R. (2007). Higher education ranking and league tables: Lessons learned from benchmarking. *Higher Education in Europe, 32*(1), 71–82.

Sadlak, J. (2007). Developments in higher education and how they stimulate the university rankings. In J. Sadlak & N. C. Liu (Eds.), *The world-class university and ranking: Aiming beyond status* (pp. 75–85). Cluj: Cluj University Press.

Sadlak, J., & Liu, N. C. (Eds.). (2007). *The world-class university and ranking: Aiming beyond status*. Cluj: Cluj University Press.

Salmi, J., & Sayoran, A. (2007). League tables as policy instruments: The political economy of accountability in tertiary education. In J. Sadlak & J. Tres (Eds.), *Higher education in the world: Accreditation for quality assurance* (pp. 79–90). Hamshire: Palgrave.

Teichler, U. (1976). *Das Dilemma der modernen Bildungsgesellschaft: Japans Hochschulen unter den Zwängen der Statuszuteilung*. Stuttgart: Klett.

Teichler, U. (1986). Strukturentwicklung des Hochschulwesens. In A. Neusel & U. Teichler (Eds.), *Hochschulentwicklung seit den sechziger Jahren: Kontinuität – Umbrüche – Dynamik?* (pp. 93–143). Weinheim: Beltz Verlag. Blickpunkt Hochschuldidaktik, Bd. 79.

Teichler, U. (1991). Towards a highly educated society. *Higher Education Policy, 4*(4), 11–20.

Teichler, U. (2007a). *Higher education systems: Conceptual frameworks, comparative perspectives, empirical findings*. Rotterdam: Sense Publishers.

Teichler, U. (2007b). Rankings: Do we need an anti-meritocratic, collectivistic and anti-dynamic higher education system? *Higher Education Digest* (Centre for Higher Education Research and Information, Open University), No. 59, Autumn 2007, Supplement, 7–8.

Usher, A., & Savino, M. (2006). *A world of difference: A global survey of university league tables*. Toronto: Educational Policy Institute.

Zhao, C. M. (2007). Building world-class universities: Some unintended impacts of university ranking. In J. Sadlak & N. C. Liu (Eds.), *The world-class university and ranking: Aiming beyond status* (pp. 321–331). Cluj: Cluj University Press.

Part II
Methodological Issues of University Rankings

Chapter 5
Ranking Indicators and Weights

Bernard Longden

5.1 University Ranking: Reliability, Consistency and Validity

> The standard market research techniques followed in data collection have raised the question: Can we assume that what works for pet food, perfume, and pesticide will also work for education? (Stella and Woodhouse 2006: 10)

5.1.1 Introduction: Positionality and Ideology

University rankings are ubiquitous and here to stay, but they are a feature of the contemporary higher education agenda. Harvey (2008: 187) reminds us that the ascendency of league tables in the higher education agenda has much to do with the 'increasing marketisation of higher education, greater mobility of students, and ultimately the recruitment of foreign students.'

The position held by stakeholders, on the worth and value of university rankings is diverse. Given the potential for a polemical position on this worth and value, it is deemed important that the ideological position for this chapter is made clear and unambiguous from the start.

A strong position is taken by Brown (2006: 38), who remains fundamentally opposed to any support of commercially produced university ranking or league tables. The basis of his criticism rests with the claim by publishers that they address matters related to quality of university teaching and research. In profoundly

B. Longden (✉)
Professor of Higher Education Policy, Liverpool Hope University, Liverpool, L16 9JD, UK
e-mail: Bernard.longden@gmail.com

J.C. Shin et al. (eds.), *University Rankings*, The Changing Academy – The Changing Academic Profession in International Comparative Perspective 3,
DOI 10.1007/978-94-007-1116-7_5, © Springer Science+Business Media B.V. 2011

challenging their proposition, he has identified four arguments to support his opposition; they are:

- Rankings are based on data and assumptions about data that are scientifically questionable
- University rankings will influence universities to produce the 'wrong' kind of higher education
- League tables reinforce the tendency to see higher education as a product to be consumed rather than an opportunity to be experienced and viewed as being 'just another commodity'
- Risk of allowing commercial considerations inexorably leads the university to a position where the market determines quality. More generally, the creation of the impression that some institutions are better than others when in a diverse, mass system there can be no one 'best university' or single view of quality. League tables indeed strengthen the market position of institutions that are already prestigious and well funded, at the expense of those that may be seeking to build reputation by attending to the needs of students and employers

Aligning to this position is to conclude that rankings misrepresent the work of universities and colleges in the interest of selling newspapers.

While this strong line of argument is attractive, there is a risk of 'tilting at windmills' which may not be so productive. In a report to the Standing Conference of Principals, my colleague Mantz Yorke concluded our report with an acknowledgement that, while they have serious limitations, it is better to work to improve them and as Lennon and McCartney say in the words of Hey Jude, '… take a bad song and make it better.'(Yorke and Longden 2005: 35).

This scepticism, which has been outlined above, remains powerful and has influenced the internal logic to this chapter and therefore this chapter reflects an ideology sympathetic to this scepticism.

5.1.2 It Starts with Events

Much of our daily routine relies on data. For example, your morning ritual may involve standing on the bathroom scales, looking at the readout, knowing that while it may not give the weight desired, it displays, accurately, numbers which are consistent and reliable. The readout from a bathroom scale is neither spurious nor idiosyncratic – it has a meaning.

The bathroom scales is a *black box* – a metaphor.

Concerns about the internal working are of limited interest as long as the readout remains accurate – accurate to a degree that is suitable for personal needs. Most of us are disinterested in finding out about how it all works, or what assumptions and adaptations have been applied.

In the physical world, we secure and gain confidence about the various pieces of physical equipment through experience. In the social world, we are dependent on trust.

An assessment about the reliability and the esteem of the person providing answers to our questions is critical to gaining that confidence.

Confidence that the metaphor of the bathroom scales is underpinned by theory is sufficient for us to entrust it.

Demands on our busy force us to place a degree of trust in the many *black boxes* we come across each day. Opening the *black box* each time, we need to be reassured that what we are dealing with s valid, reliable, and consistent, which is not a realistic possibility.

Life would become impossible without some degree of confidence and trust in the validity of the information that the physical or social world generates.

The concept of trust becomes an integral part of the human condition; without it, our every experience would become a series of hypothesis open to rigorous testing before decision could be made about the experience.

5.1.3 Locating a 'Black Box'

The thesis, underpinning this chapter, is that compilers of university rankings rely on us not opening the *black box*; not to question the interrelationships between the various parts within the *black box*; and not to question the construction of the various elements within the *black box*.

Constructing the university ranking relies on a conspiracy between the compilers and the wider community that the *black box* remains unchallenged. The mathematical and statistical complexity embedded within the *black box* would demand too much valuable time and effort to unpick, leaving the reader of the rankings depend on a trust in those who construct the rankings; after all, the reader is probably only interested in the final column – the aggregate sum of the other columns shown in the table.

The final ranking index provides some form of objective measure – a measure that, in some real sense, maps an underlying set of characteristics that, when aggregated in a particular way, provides a ranking.

Few are prepared to dig beneath the top line index to find out how the final measure was achieved – how the *black box* actually works. Compilers remain confident that it is unlikely that serious criticism will come from an analysis of the content of the *black box* because of the 'challenging mathematical and statistical complexity' that inevitably would ensue.

By negating a willingness to open the *black box* and peer inside, we have handed over to the compilers a consent that what they have done is acceptable – that they have provided a reliable, consistent, and valid processes that measure or reflect the 'worth' of the universities ranked.

So, to open the *black box* requires effort and a sense of a critical commitment to understand how the mechanism works. For most of the time we simply want a quick answer to the question 'which is the best university?' as if such a deceptively simple question can mask the complexity beneath.

Table 5.1 Problems intrinsic in designing university rankings

1. Become an end in themselves protected from critical scrutiny
2. National, institutional and program diversity
3. National and linguistic diversity
4. Partial coverage of purposes and stakeholders
5. Problems of aggregation and weightings
6. Reputational rankings
7. Produced context-free judgements
8. Undermine universal improvements
9. Reduce scope for innovations in strategy, curriculum, pedagogy and research

Other chapters provide a more detailed discussion on the use and abuse of university rankings; however, it is appropriate at this point to be reminded of some of the main problems associated with the process of ranking universities.

Table 5.1 provides a list of some of the intrinsic problems that university rankings generate. In this chapter, the focus will be on the methodological issues that arise during the creating of a university ranking index, although some of the other problems identified in Table 5.1 may be addressed in passing.

There remains a challenge that when the *black box* is opened is it possible to understand the processes deployed by the compilers? Do compilers ensure satisfactory levels of transparency in describing their methodologies? Is there a commonality to the methodologies that different compilers employ in constructing the ranking index? How visible is the internal mechanism of the *black box*?

I propose to focus on these questions to gain an insight into the problems associated with aggregation and weightings of performance indicators in Table 5.1. Along the way, I will briefly address other issues listed in the table but the main objective will be the focus on the key confidence measures or validity, reliability and consistency as these remain the cornerstones of the trust that is given to those compiling the university rankings.

5.2 Critical Steps and Economical Truths

A university ranking index provides an end point for the user by a process of consolidating a large data set, a single index that in some ways represents the ‘university’. This simple statement exposes the facile nature of the process. How can the activity of a university be reduced to a numerical value? Anyone who works within a university setting knows only too well that within the one institution, there are pockets of high quality and pockets that are of concern to the institution. Providing a single measure betrays the complexity of the institution. Unlike a car manufacturer where there is a product line to measure, universities have different aims from each other and therefore comparisons that fail to take note of the differing ‘missions’ fail to make sensible comparisons.

Clarke (2002, 2004) describes two common criticisms relating specifically to *US News* and *World Report* and the methodology used to rank colleges and universities. First, the compilers constantly change the formula they use to create the rankings and thus make the interpretation of yearly shifts in a university/college ranking, in terms of academic quality, impossible. Second, the score used to assign schools to ranks is overly precise, creating a vertical column where a group might more properly exist.

In this section I plan to tease out aspects of the mechanisms within the *black box* to ensure that there is an understanding of the techniques deployed. These techniques need to be understood so that assumptions implicit within the process can be appreciated and create a more transparent methodology capable of evaluation by the user. There are a number of specific elements within the methodological *black box* that will be considered:

- Selection of indicators to produce the final ranking index, issues surrounding the way in which indicators are added together
- The relative weighting that is applied to the various indicators deployed to create the final ranking index
- Management of missing data
- Statistical differences between the ranking indices that emerge

Seven steps can be identified clearly in the process of creating a university ranking; consideration of each of those stages follows.

5.3 Steps Towards Creating a University Ranking Index

The problem with ranking concerns the practice not the principle. (Altbach 2006).

How is it possible to accurately reduce a university performance to a single index? A university is a cauldron of beliefs, values, and actions and the proposition that it could be possible to distil this all down to a single index remains, for me, a challenge and a fear that there may be more reliance on alchemy than on logic.

However, it would be inaccurate to suggest that compilers involved in preparing and publishing university rankings seek to keep the box tightly closed. Quite the reverse, most seek to provide the reader with a very detailed account of how they compile the ranking indices (Morse and Flanigan 2006; O'Leary et al. 2006; MacLeod and Hiely-Rayner 2007, 2008, 2009a). Recently, the *Times Higher Education* has been at pains to make adjustments to the methodology they use in calculating the 'world rankings' (Baty 2009).

In considering the stages necessary in producing university rankings, it is possible to identify key processes that all compilers appear to adopt. For the purpose of this chapter, I propose to explore each of these stages in some detail. The starting point of the process is the measurement of an event that relate directly to the university activity. The measurement or performance indicator (PI) – is ubiquitous, often it is invisible, which helps define the institution. When aggregated with other

Table 5.2 Critical steps that pose potential difficulties when creating a university ranking

1. Clarifying reason for ranking
2. Selecting suitable metrics – performance indicators (PI)
3. Collecting data – metrics
4. Adaptation of PIs into a scale
5. Standardising measures prior to aggregating
6. Weighting PIs prior to aggregating
7. Creating a single index reflecting a university

measures, it can provide a numerical shorthand for key characteristics about the university.

The steps deployed need to be identified and confidence need to be secured so that assumptions, adaptations and definitions are fit for purpose. Using the critical steps as a guide, it is possible to show points where potential difficulties can occur and are often overlooked or ignored when providing a narrative on how rankings are created.

Teasing out these critical steps provides a means by which those elements of the process that are vulnerable to mystification and obscuration can be considered in detail (Table 5.2).

5.3.1 Clarifying Reason for Creating University Ranking

Altruism is unlikely to be the justification advanced by a publishing company engaged in producing a university ranking table. The reason why publishers involved in this genre of publishing retain their involvement is simply down to the money they generate from the final product – advertising revenue, purchases of the final ranking book and other forms of endorsement. The really great thing for the publisher is that once the template for the production of the tables has been established each year, a new target population is ready to buy their product.

Two main types of university rankings are evident in the commercial world of rankings. The audience for the two types of university rankings is distinct and different but the methodology adopted by compilers to create the rankings is very similar.

5.3.1.1 Type 1: Undergraduate Experience: Teaching

US News and *World Report* (*USA*), *Maclean's* (Canada), *The Guardian* (*UK*) and *The Times* (UK) all have as their target audience the potential undergraduate student market. The common feature of all these and other similar publications is the production of a ranking that, it claims, reflects the quality of teaching and learning within higher education institutions. The measures that are used to reflect this quality index are those that relate directly or indirectly to the undergraduate experience.

The complexity of the data that is required to produce comparisons is such that when coupled with the diversity of provision across countries, the university rankings tend to be specific to a country. Attempts at providing the global rankings for teaching and learning have remained elusive so far. The main reason for this must relate the need to secure a common set of definitions for the measures employed in the creation of the rankings. Given the diversity of provision this remains an obstacle.

The nature of measures typically associated with domestic university rankings would be student staff ratios, spend per student FTE, student satisfaction measures. Even within one country, the diversity of measures or PIs used in the calculations signal that there is no common agreement on the definition of what constitutes high quality provision.

5.3.1.2 Type 2: Postgraduate Research Ranking

Both Shanghai Jiao Tong University Academic *Ranking of World Universities* (SJTU *ARWU*) and the Times Higher Education (Quacquarelli Symonds Ltd) *World University Rankings* have exclusively focused on rankings in relationship to quality of research provision. The target audience could be considered to be potential funding sources and potential academic researchers. Typically, research ranking measures include articles published; papers cited; research student numbers, prestigious awards for research secured, etc.

The critique that follows applies to both types of ranking however the source and nature of the performance indicators used in the calculation will be substantially different, not only between the primary purpose of the ranking but also between the different publishers engaged in producing the rankings.

For example, a focus on the student market will focus on the nature of the learning environment, and facilities and resources available for the student, whereas a research focus will be on the track record for research secured by the university, and the research facilities available often coupled with peer esteem of the research status of the university.

The following section will draw, as appropriate, on both types of rankings.

5.3.2 Selecting Suitable Metrics: Performance Indicators

5.3.2.1 Performance Indicators

It is therefore not surprising that the Performance Indicator (PI) has helped form the landscape of higher education, providing a critical measure to help answer the question: How do I know what I am achieving? (Cave et al. 1997).

As such, performance indicators (PIs) are designed to provide quantifiable measurements which, having been agreed in advance, reflect the relative success of an organisation (Longden 2008). However, what is selected for measurement

is governed by the nature of the organisation and is political – political with a small '*p*'. Who decides to record student entry qualifications rather than student socio-economic background exposes a particular interest in the characteristics of students in higher education.

PIs are usually seen as numerical measurements of achievement that are easy to collect, interpret and use, with the emphasis on 'easy to collect'. In theory, PIs can only be derived from things over which direct control can be exerted leading to achieving an outcome of the measure. It is not surprising that PIs are of interest to a wide range of bodies, ranging from federal and local governments agencies, through to universities and colleges themselves, and, ever increasingly, students.

With the student market in mind, compilers of university rankings would claim that they have attempted to simplify a complex set of PIs measures by aggregating them to form a single index, sorted in order thus producing the university ranking.

The claim is made by compilers that university rankings 'help' potential students and their parents to reduce the mass of information about the universities and in doing so, they claim they are assisting in the decision making and enabling students to come to the conclusion about the 'right university to attend'.

With over 4,200 accredited universities in the USA and about 130 in the UK, for example, it is clear that the task facing a prospective student in selecting the 'right' university is a daunting one not only for the prospective students but also for concerned and interested parents. University rankings clearly service a need.

5.3.2.2 Proxy Measures

Given the origins of the data, it would not be surprising that compilers often require data that is not provided in the direct measurements provided by the sources discussed above. Teaching quality is one such measure that is deceptively simple and would be expected to be easily available but is neither. Compilers are forced to consider other ways of achieving the measure. In the UK, the measure is derived as a proxy measure from the *National Student Survey*[1] (NSS), while the *US News and World Report* in the USA derives the measure from a dubious logical connection between '*alumni giving and satisfaction*'. It could be argued that each PI should be scrutinised to ensure that what it measures and what it purports to represent in a ranking are sufficiently close to be acceptable.

5.3.2.3 From Judgement to Number: What Is Regarded as Important

Each event of observing the world evokes a judgment of what we decide to record about the event, and what particular part of the experience is important at that

[1]National Student Survey is conducted in the UK as a statutory requirement on all higher education providers to ensure that over 60% of their final year students contribute to the web based survey.

moment in time. Experience is not naturally coded as a set of numbers; we frequently impose a number at a later date and time when describing the event.

Within a commercial setting, it is possible to move from judgment to a numerical measure with greater ease than within the education setting, where it is often difficult, maybe impossible, to make hard measures from a socially constructed experience.

Graham and Thompson (2001) argue that most prospective students and their parents require reliable comparable information on the most important outcome of a college education, namely:

- What have I gained by way of learning from this experience–*learning outcomes*?
- Has the total experience rated highly on the *student satisfaction* index?
- Have I worked sufficiently effectively that I gained a certificate that will be acknowledged by others as a measure of success – *graduation*?

Interestingly, these apparently simple measures are dependent on proxy measures and rely on a simple relationship.

$$\text{Good student} + \text{Good faculty} = \text{Good university}$$

With this simple model, many compilers have set about to construct a university ranking that then teases out measures about faculty and students to help construct the metric of 'good university'. Frequently, compilers make use of measures relating to student entry qualifications, faculty qualification, i.e., percentage of doctoral staff, all of which are proxy metrics for measure that is more elusive to grasp- hold. Graham and Thompson suggest that:

> … [it is] like measuring the quality of a restaurant by calculating how much it paid for silverware and food; not completely useless, but pretty far from ideal. (Graham and Thompson 2001)

Despite best efforts, data are complied and reported according to value judgements that are embedded in the methodologies for collection and reporting; some of these value judgements are explicit, some implicit.

> Hardly a week goes by without another league table measuring university performance … . Of course none of these tables are the same; they all use different statistical measures and weightings to reach their judgments. While some focus on teaching quality, others emphasise research or take greater account of students' entry qualifications, the money spent by institutions, the proportion of students who get a 2:1 or the percentage who get a graduate job. Not only do these measures vary between papers, they also vary from year to year. So, while government teaching inspection scores might be important one year, it could be the level of library expenditure the following year. (Morris 2005)

5.3.2.4 Outputs, Inputs and Process

A helpful means of differentiating the different measures that are available in creating rankings, be it for ranking universities or subject within universities or research generated by universities, is to classify the measures into the three types of PI – outputs, inputs and processes.

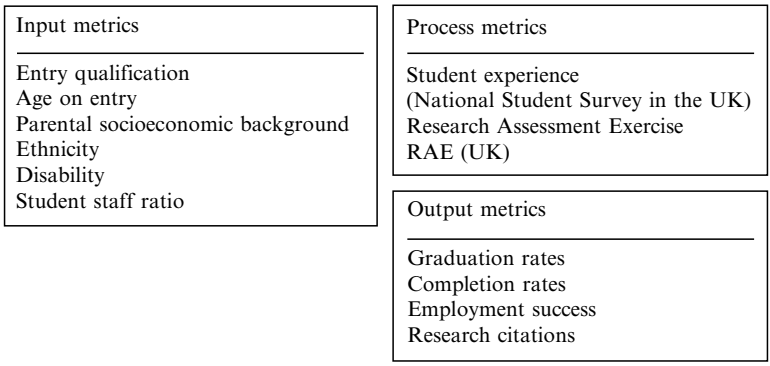

Fig. 5.1 Broad categories and types of metric

Input measures might include qualifications on entry, student staff ratios, resource expenditures, while output measures might include successful completion rates, employment characteristics and degree success rates, citations or published articles. Processes, while being harder to measure, relate to the quality measures for research or teaching; for example, in the UK, the National Student Survey (shown as a process measure in Fig. 5.1) provides a measure of the student experience; it does not provide a measure of the student learning.

The point is well made by Richardson (2008: 18) that few compilers in the UK make any clear distinction between the three types of metrics shown in Fig. 5.1 and that the over-emphasis on input is to the detriment of the overall ranking methodology. Richardson, citing work by Pascarella and Terenzini in 2005, suggests that there is little evidence to support input measures as they 'have only an inconsequential or trivial relationship with student learning and achievement'.

5.3.3 Collecting Data

Both UK and US compilers attempt to make maximum use of authoritative data. As will be discussed later in this chapter, confidence in how authoritative the data may be has been challenged in the USA. In the UK, the data used by most compilers has been collected through an agency of central government.

The creation of a ranking index depends on the selection of data, data originating, as I have argued, from events of different forms and complexity within the life of the university. Compilers of university ranking indices draw on this data to help construct the ranking. Three types of data are available for use in the ranking tables,

- Primary data generated by the university itself
- Survey data generated by the compilers
- Data collected from independent third parties

Primary data produced by universities for both internal and external purposes has been subjected to external audit. In the UK, there is a statutory responsibility placed on all publicly funded institutions to provide data to Government Agencies to support the financial investment made. The data is subject to tight auditing which requires clear precise definitions of the data used, specific dates around which data is collected, recorded, and transmitted. It is this audited data which, if released to the commercial compilers, is used to create the ranking index. In the USA, data provided to the commercial compilers is provided directly by the institution. The audit function is absent. Usher and Savino note that in the USA:

> there is no guarantee that institutions will actually report the data to the rankers on a consistent basis , as all have a clear incentive to manipulate data in a manner which will benefit them (Usher and Savino 2007: 26)

This prophetic insight by Usher and Savino has materialised itself in revelations from two colleges in the USA. In both cases, the risk implicit in self-regulation associated with submitting unaudited data to the compilers of ranking tables is evident.

The first case is that of Clemson University, which claims to be one of America's top public research universities, located in South Carolina. A conference presentation[2] by a member of Clemson University staff exposed the practice of managing data to secure an improvement year-on-year on the rankings. Among the steps reportedly alleged by Watt, who until 2006 headed Clemson's institutional research office were that Clemson:

- Manipulated class sizes
- Artificially boosted faculty salary data
- Gave rival schools low grades, which counts for 25% of the score in *US News* and *World Report's* peer reputation survey

Irrespective of the actual truth in the claim and counter claim, the weakness is there for all to see. In essence, Clemson University submitted data that ensured the University rankings moved from 38th to 22nd position in *U.S. News*' ranking of public research universities from 2001 to 2008.

> The easiest moves … revolved around class size: Clemson has significantly increased the proportion of its classes with fewer than 20 students, one key U.S. News indicator of a strong student experience. [I]t has focused … on trying to bump sections with 20 and 5 students down to 18 or 19, but letting a class with 55 rise to 70. 'Two or three students here and there, what a difference it can make', Watts [Clemson administrator responsible for managing the US News submission] said. It's manipulation around the edges. (Lederman 2009a)

By creatively managing the class size data in this way, it was possible to ensure that Clemson University PI for student data was maximised for the faculty resources element of the index.

[2]Title of AIR presentation Strategic Project Management: Lessons from a Research University's Successful Drive to Move Up in the US News Rankings by Catherine E. Watt – Director of the Alliance for Research on Higher Education, and Nancy T James – Research Analyst III, Clemson University.

How widespread an occurrence of this 'creative management' of PIs is hard to assess – that it occurs at all is no longer in doubt.

The second case relates to self reporting of data submitted to *US News* and *World Report* by the University of Southern California (USC). USC claimed that 30 of its professors were members of the prestigious National Academy of Engineering; on its Web site, the engineering school went even further by listing 34 such professors (Lederman 2009b; Shea 2009). Further investigation provided evidence that the claim was a substantial over estimate of the actual number of 22.

Clemson and USC are not the only institutions susceptible to the pressures to managing their ranking status. Ehrenberg (2003) in the USA and Watson (2008) in the UK have written on this topic.

Survey data may be developed specifically by the compilers (as in the case of *US News* and *World Report* which incorporates a peer reviewed surveys (see Table 5.3) contributing 25% to the final ranking score) or may be extracts from national surveys as in the case of the UK's *National Student Survey* (NSS).

> The National Student Survey, which measures student satisfaction, will be factored into the rankings for the first time. As a result, figures that represent the subjective sentiments of those who are willing to fill in the forms will be turned into seemingly objective measures of the worth of higher education institutions (Ryan 2009).

The NSS data was developed to provide a measure of the quality of the student experience of higher education in the UK. It remains a statutory responsibility for each higher education institution in receipt of public funds to provide a minimum survey response rate[3] anonymously completed by its final year students. While there have been claims of influencing the student opinion about their experience (Newman 2008), the extent is limited.

The use of surveys developed by compilers remains an important component of the *US News* and *World Report's* methodology. The limitations of this approach have been highlighted by the Clemson clarification that the President's completion of the Peer review submission exposes the difficulties in being altruistic when self reporting.

> … Clemson officials, in filling out the reputational survey form for presidents, rate "all programs other than Clemson below average," to make the university look better. "And I'm confident my president is not the only one who does that" (Lederman 2009a)

The Times rankings in 2008 introduced the approach well established by *US News* and *World Report* by seeking the opinions of Heads of secondary schools and from university academics about where the highest-quality undergraduate provision was located. It was probable that *The Times* had borrowed the approach from *US News* and *World Report* – wisely, this type of data collection was dropped for the following year calculations!

Independent third party data, usually from administrative source such as government or grant making bodies, are generally regarded as '*the gold standard of*

[3]NSS – the minimum response rate for 2009 was 60%.

Table 5.3 US News and World Report Weights for national universities and liberal arts colleges

Metric		Sub-factors		Overall contribution – weighting	
		National universities and liberal arts colleges (%)	Universities master's and baccalaureate colleges (%)	National universities and liberal arts colleges (%)	Universities master's and baccalaureate colleges (%)
Peer assessment	Peer assessment survey	100	100	25	25
Student selectivity	Acceptance rate	10	10	15	15
	High school class top 10%	40	0		
	High school class top 25%	0	40		
	SAT/ACT scores	50	50		
Faculty resources	Faculty salary	35	35	20	20
	Staff qualifications	15	15		
	Full-time faculty	5	5		
	Student faculty ratio	5	5		
	Class size 1–19 students	30	30		
	Class size 50+ students	10	10		
Graduation and retention rate	Average graduation rate	80	80	20	25
	Average freshman retention	20	20		
Financial resources		100	100	10	10
Alumni giving	Average alumni giving	100	100	5	5
Graduation rate	Graduation rate performance	100	0	5	0

Source: Morse and Flanigan (2009)

comparative data since they are, at least theoretically, both accurate and impartial' (Usher and Savino 2007: 26). In their survey of global ranking systems, Usher and Savino make the point that while accuracy and impartiality might be positive aspects of their contribution, they are really administrative by-products of data collected for other purposes, opening up the potential for using a measure out of its original context.

The plurality in use of data sources varies considerably even within one country where two or more commercial compilers operate. This should raise concerns among those who depend on the rankings as it implies inconsistency in the principles adopted by each complier as to where the data comes from.

The '*Times* ' (Times online 2009), for example, describes the 2010 tables data sources as:

> All sources of the raw data used in the table are in the public domain. The National Student Survey (NSS) was the source of the Student Satisfaction data. … The information regarding Research Quality was sourced from the 2008 Research Assessment Exercise … Staffing data supplied by HESA were also used to evaluate the extent to which the research ratings related to total academic staff. … Entry Standards, Student-Staff Ratios, Services & Facilities Spend, Completion, Good Honours and Graduate Prospects data were supplied by the Higher Education Statistics Agency (HESA) which provides a system of data collection, analysis, and dissemination in relation to higher education in the whole of the United Kingdom. The original sources of data for these measures are data returns made by the universities themselves to HESA (Times online 2009).

The Guardian (MacLeod and Hiely-Rayner 2009b), drawing on similar sources but from a very different set of performance indicators, have used the HESA and NSS data in their 2009 calculation of a ranking metric.

> The rankings are compiled from the most recent figures available - official 2006-07 returns from universities and higher education colleges to the Higher Education Statistics Agency (Hesa). They also incorporate data from the National Student Survey (NSS) 2007, published by the Higher Education Funding Council for England (Macleod and Hiely-Rayner 2009a).

The pretence at being objective and quasi scientific has been discussed by Stella and Woodhouse (2006: 6), where they suggest that there are generally two broad data types: data provided by institutions and data derived from expert opinion, both giving an illusion of being 'scientific' and by inference, beyond reproach.

> Most rankings rely on two types of data - data given by institutions that is accepted, often without a reliable validation process, and data obtained from opinion polls in the name of 'expert opinion'. With both components on shaky grounds, the use by the media groups of complex formulae with weights and indicators only helps to project a pseudo image of being 'scientific' to outcomes that may be statistically irrelevant (Stella and Woodhouse 2006: 10).

There is a clear necessity for data to be managed within an institution in ways that take account of the uses to which they are, and might be, put. The importance of data definition and management for how the data is returned to the compilers of the rankings is such that, at all levels within an institution, staff are aware of the ways in which what they do, and how it is recorded, could have a significant impact on their futures.

A sudden decline in an institution's position in the rankings, which might derive simply from particular choices in collating and reporting data, could for instance have a sharp adverse effect in the international market for higher education. It matters greatly how an institution presents truths.

The process of creating a university ranking index starts within the university and the events that constitute that university. These events are various, complex, and frequently invisible to the casual viewer; some of the events are captured, nonetheless, for different reasons by the university. The reasons can be various too; faculty management requirements such as class lists, assignment submissions etc.; internal management of the university to ensure quality standards are maintained or facilities are supported; and external statutory requirements such as those required in the UK to support the funding model used by the Higher Education Funding Council for England (HEFCE) to distribute block teaching funds. The list of events and therefore data is substantial. From this mass of data, compilers select certain items of data to include in the ranking methodology.

The apparent simple act of capturing data brings with it intrinsic difficulties. To a university outsider, the simple event of counting the number of students on a program would not appear too challenging. However, those involved in data collection are only too well aware that data collection brings with it a set of ever expanding definitions. The quotation from HESA (2009) illustrates the increasing complexity of the data definition required by universities. Data submitted in the UK is provided to both the Higher Education Funding Council for England (HEFCE) – which is an aggregate data set – and to the Higher Education Statistics Agency (HESA), where an individual's student record is submitted. Considerable pressure is placed on universities to ensure that data quality is high; both HESA and HEFCE have sophisticated data audit systems operating to ensure consistent, accurate data is provided. As a final pressure on universities, HEFCE operate a data audit on institutions to maximise data quality.

Subsets of the data are released by HESA, after data protection agreements for each data request has been agreed (or not as the case may be), to UK compilers of university rankings.

It is important to note at this point that the data provided in the England to HESA was originally provided to support funding claims against HEFCE; the data was not collected to assist compilers with the production of their university rankings.

> The HESA session population has been derived from the HESA Student Record. It includes all higher education and further education student instances active at a reporting institution at any point in the reporting period 1 August to 31 July except: dormant students (those who have ceased studying but have not formally de-registered) incoming visiting and exchange students. Students where the whole of the programme of study is outside of the UK, and from 2007/08: students on sabbatical.
>
> Incoming visiting and exchange students are excluded from the session population in order to avoid an element of double-counting with both outgoing and incoming students being included. The HESA session population forms the basis for counts of full-time equivalent (FTE) student instances (HESA 2009).

In the extract above from the HESA guide to higher education institutions for submission of data in 2009, the complexity and need for very precise definitions

is evident. This is partly why comparative data is difficult to obtain. Where can there be confidence in the precise mapping of data across educational jurisdictions? What, for example, is implied by the deceptively simple term "de registration"?

5.3.4 Adaptation of PIs into a Scale

The selection of metrics for inclusion reflects the objectives that are to be achieved by the ranking process. In *US News*, *The Guardian,* and *Times,* a measure used in the overall ranking calculation is a financial one. For example, faculty compensation in *US News* is the average faculty pay and benefits adjusted for regional differences in cost of living. In the case of *The Guardian* and *Times,* the data is taken directly from the HESA finance return and is a ratio of spend per student full time equivalent (FTE). In both examples from the USA or UK, the final measure is $ per faculty or £ per student FTE. It would not be possible to incorporate these values directly into any calculation of ranking without an adaption.

Richardson (2008: 20) notes that the process of adapting the data, in readiness, for aggregation is frequently termed, incorrectly (in the strict statistical sense), as normalisation. It encompasses the process of adapting the data to reflect adjustments necessary when dealing with institutional size or institutional subject/discipline composition. It is acknowledged in the UK that the national funding model positively advantages institutions with significant medical schools when spend per student is considered (evidence from the USA and Australia suggest a similar effect occurs there too). Compilers, in their attempt to deal with this distortion, apply a modification to the metric to account for this 'distortion'.

Data used by the Guardian's 2009 guide for spend-per-student studying. Sociology indicates that the range of data is from £407.99 to £3,243.49. This is calculated from the amount of money that an institution spends providing a subject (not including the costs of academic staff, since these are already counted in the staff-student ratio) adjusted to account for the variation in volume of students learning the subject. Figure 5.2 also includes the money the institution spends on central academic services, and per student FTE.

In discussion with a compiler of the Guardian's table, it became clear that while the actual data was incorporated directly into the calculation, for display and publication purposes, and to avoid issues related to publishing actual data in the table, the data was transformed to a single 1–10 scale.

The adaptation of data into a scale is frequently used in the methodology adopted by *The Times*. The construction of the scale is arbitrary and not based on any theoretical analysis. The assumption is that the scale is linear; but there is no justification for that assumption. Why not log linear or inverse or sigmoid?

Either the ranking lends itself to a scale of 0–100 or to a band to which numerical values can be applied. Whichever detailed process is used, the final product is a numerical value for the PI which can then be used directly in producing the final index.

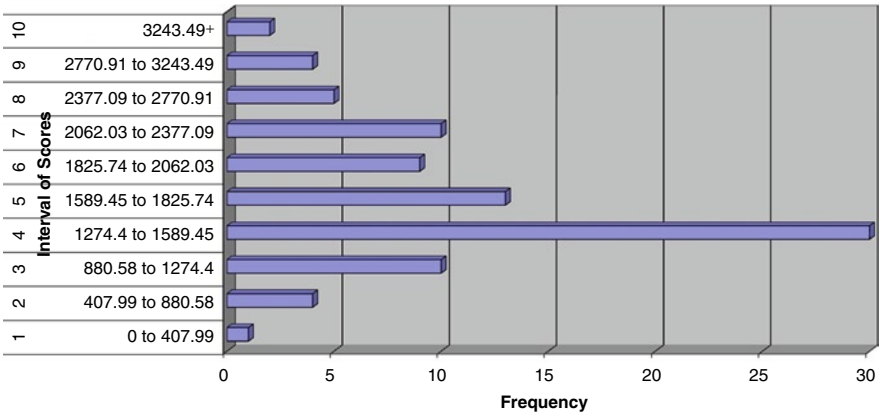

Fig. 5.2 Distribution of spend per student in sociology (Source: MacLeod and Hiely-Rayner, 2007)

It is interesting to note that the instructions provided by *US News* on the Faculty Compensation measure (see Morse 2009 for a detailed definition of the measure) hint at the importance of the metric in the overall calculation of the ranking score. It notes that "…*higher average faculty salaries score better in the ranking model*" (Morse 2009). The model used by *US News and World Report* in the construction of their college ranking tables rewards institutional expenditure over any other institutional measure used in the calculation of the final ranking score.

5.3.5 Standardising Measures Prior to Aggregating

A relatively inconsequential paper published in 1955 provides a challenge to anyone attempting to add different types and sources of data together (Richmond 1968: 182). The paper published in the *Journal of incorporated Associations of Assistant Masters in Secondary Schools* describes a simple scenario where test scores for ten subjects were set out in a table for ten children. The scores were added together to provide an aggregate score for each individual. From the total, it becomes clear who should be labelled '*top of the class*'. However, on closer scrutiny it becomes clear that each subject has used a different range of marks – some used the whole scale from 0 through 100, while others use a scale from 30 through 65. When the variability in use of the scale is incorporated into the calculation, the rank order in the class is reversed.

The implications for those engaged in aggregating data from different sources and from different distributions are both significant and important.

Adding scores together is simple but it can only be acceptable when the scores have been adjusted so that the distribution and range of the scores conforms to a specific format. In the next section, consideration is given to the nature of that format and the conditions that need to be achieved before confidence can be given to the outcome.

5.3.5.1 Transformations

To ensure that when adding the two data sets together the same 'measures' are added together, it is necessary to treat the two data sets by stretching them so that they conform to common statistical measures where the mean value is zero and the standard deviation is 1. This transformed distribution thus created is referred to as the 'z-score'. The 'z-score' transformation requires some basic understanding of statistics.

The problem of adding two PI scores is best illustrated in Table 5.4 based on data from 26 universities. A closer statistical examination of each of the PIs shown in Table 5.4 indicates that the range of data for *PI* A ranges from 90 to 20 while that for *PI* B ranges from 62 to 43. When the mean scores are compared, a further

Table 5.4 University rank order created from two performance indictors

University	*PI* A	*PI* B	Sum (*PI* A + *PI* B)	Rank order
U_001	*85*	*54*	139	**1**
U_002	*85*	*50*	135	**2**
U_003	*90*	*44*	134	**3**
U_004	*74*	*51*	125	**4**
U_005	*78*	*46*	124	**5**
U_006	*76*	*44*	120	**6**
U_007	*70*	*50*	120	**6**
U_008	*64*	*53*	117	**8**
U_009	*62*	*55*	117	**8**
U_010	*64*	*52*	116	**10**
U_011	*60*	*56*	116	**10**
U_012	*62*	*52*	114	**12**
U_013	*64*	*45*	109	**13**
U_014	*45*	*61*	106	**14**
U_015	*51*	*54*	105	**15**
U_016	*47*	*57*	104	**16**
U_017	*50*	*51*	101	**17**
U_018	*54*	*47*	101	**17**
U_019	*34*	*62*	96	**19**
U_020	*28*	*57*	85	**20**
U_021	*42*	*43*	85	**20**
U_022	*35*	*50*	85	**20**
U_023	*30*	*54*	84	**23**
U_024	*35*	*49*	84	**23**
U_025	*24*	*50*	74	**25**
U_026	*20*	*44*	64	**26**
Min	*20*	*43*		
Max	*90*	*62*		
Sum	*1429*	*1331*		
Mean	*54.96*	*51.19*		
Number	*26*	*26*		

difference becomes evident. (*PI* A = 54.96 and *PI* B = 51.19). This illustrates clearly that the two data sets are different and that any attempt at aggregating each of the individual PIs together would present a problem.

5.3.5.2 *z*-Scores: Calculating Standardised PIs

The '*z-score*' provides two important characteristics about performance indicators

- The relative position of the PI measure relative to the mean
- The distance from the mean

Negative 'z-scores' indicate PI measures below the mean; positive *z*-scores indicate PI measures above the mean. 'z-scores' with a larger absolute value are further away from the mean from *z*-scores that those that are smaller in absolute value (−2.30 is further from the mean than .40).

The calculation of a 'z-score' can be considered in two stages.

The first stage introduces the concept of spread of data around the mean. The mean value ($\overline{x}$ described as bar x) is calculated from a summation of the all the university scores divided by the number of scores contributing to the total score. The spread of data for each university from the mean is aggregated, i.e., $x - \overline{x}$, where x is the individual value for each university and $\overline{x}$ is the mean for all the universities.

$$z = \frac{x - \overline{x}}{SD}$$

From Table 5.4, the mean value for all universities for *PI* A is 54.96; the value specifically for university 004 is 74, giving a difference from the mean of 19.04 (when each individual measure from the *PI* A mean is aggregated the net result is a mean of the spread or variance value of zero).

The second phase of standardisation involves a calculation of the standard deviation of the university PIs. To achieve this, each variance score is squared and aggregated and then divided by the number of universities contributing. The formula is shown below:

$$s = \sqrt{\frac{r_1^2 + r_2^2 + \cdots + r_n^2}{n-1}} = \sqrt{\frac{\sum \left(x_i - \overline{x}\right)^2}{n-1}}$$

When this is carried out for each university score both positive and negative numbers result. Statisticians frequently use the device of squaring a number to remove the problem of negative values. The standard deviation for *PI* A is 20.10.

The amended Table 5.5 (based on Table 5.4) has been extended to show the calculations of the *z-score*s for each university together with a comparison of the difference between the rank based on the raw data from the ranking derived *z-score transformations*.

Table 5.5 Comparison between rank order created from raw scores and transformation score ('*z* score')

	RAW				*z* score				
University	*PI* A	*PI* B	Sum (*PI* A + *PI* B)	Rank	*PI* A	*PI* B	Sum *z*(*PI* A + *PI* B)	Rank	Differences
U_001	85	54	139	**1**	1.49	0.11	1.60	**1**	***0***
U_002	85	50	135	**2**	1.49	–0.05	1.45	**3**	***–1***
U_003	90	44	134	**3**	1.74	–0.28	1.47	**2**	***1***
U_004	74	51	125	**4**	0.95	–0.01	0.94	**5**	***–1***
U_005	78	46	124	**5**	1.15	–0.20	0.95	**4**	***1***
U_006	76	44	120	**6**	1.05	–0.28	0.77	**6**	***0***
U_007	70	50	120	**6**	0.75	–0.05	0.70	**7**	***–1***
U_008	64	53	117	**8**	0.45	0.07	0.52	**8**	***0***
U_009	62	55	117	**8**	0.35	0.15	0.50	**9**	***–1***
U_010	64	52	116	**10**	0.45	0.03	0.48	**10**	***0***
U_011	60	56	116	**10**	0.25	0.18	0.44	**11**	***–1***
U_012	62	52	114	**12**	0.35	0.03	0.38	**12**	***0***
U_013	64	45	109	**13**	0.45	–0.24	0.21	**13**	***0***
U_014	45	61	106	**14**	–0.50	0.38	–0.12	**15**	***–1***
U_015	51	54	105	**15**	–0.20	0.11	–0.09	**14**	***1***
U_016	47	57	104	**16**	–0.40	0.22	–0.17	**16**	***0***
U_017	50	51	101	**17**	–0.25	–0.01	–0.25	**18**	***–1***
U_018	54	47	101	**17**	–0.05	–0.16	–0.21	**17**	***0***
U_019	34	62	96	**19**	–1.04	0.42	–0.63	**19**	***0***
U_020	28	57	85	**20**	–1.34	0.22	–1.12	**23**	***–3***
U_021	42	43	85	**20**	–0.64	–0.32	–0.96	**20**	***0***
U_022	35	50	85	**20**	–0.99	–0.05	–1.04	**21**	***–1***
U_023	30	54	84	**23**	–1.24	0.11	–1.13	**24**	***–1***
U_024	35	49	84	**23**	–0.99	–0.08	–1.08	**22**	***1***
U_025	24	50	74	**25**	–1.54	–0.05	–1.59	**25**	***0***
U_026	20	44	64	**26**	–1.74	–0.28	–2.02	**26**	***0***
Sum	1429	1331			0.00	0.00			
Mean	54.96	51.9			0.00	0.00			
Number	26	26			26	26			
SD	20.10	5.15							

5.3.5.3 Impact

To illustrate the impact of the two ranking methodologies, Fig. 5.3 provides a graphic for the difference between the raw ranking position and the position based on the 'z-score' transformation. For some universities, the impact is insignificant; for example, U_001 is un-affected by the transformation and remains at the top of the rankings, whereas U_020 based on the raw score ranking was 20th, yet when based on the 'z-score' transformation was adjusted to the 23rd position. The impact of applying a 'z-score' transformation to the raw data shows up very clearly in the apparent random changes that occur between the universities.

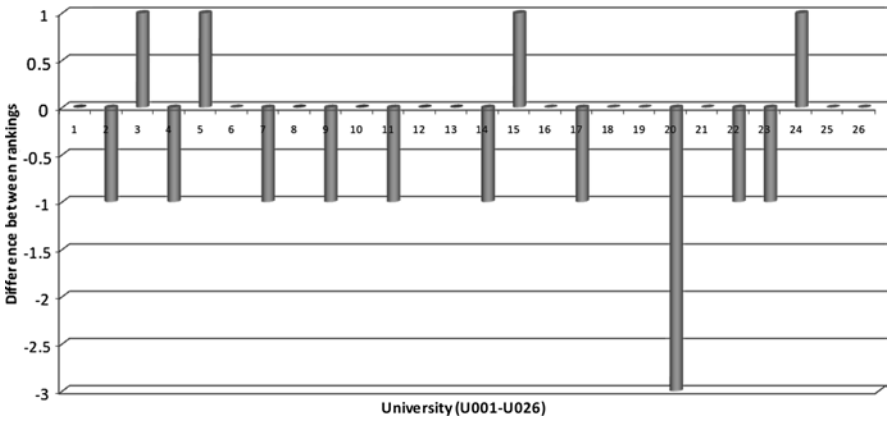

Fig. 5.3 Impact of applying a transformational (*z* score) when aggregating two PIs

In summary, the '*z*-score' provides a measure of the number of standard deviations (SD) each *PI* measure is away from the mean. For example, a '*z*-score' of 1.3 means that the PI was 1.3 SDs above the mean, whereas a *z*-score of −.70 means that the PI is .70 SDs below the mean and a *z*-score of 0.00 indicates a PI exactly the same as the mean.

By re-calculating *z*-scores for each PI, we have essentially re-scaled, or re-numbered the scores. In other words, we have essentially changed the scores from their original values to new values that are directly interpretable. Because *z*-scores are linear transformations, we have not changed the shape of the distribution.

For a detailed explanation of the underlying theory associated with '*z* score' transformations it is suggested that you refer to appendix A of Richardson's report (2008: 6). Standardisation refers to the process of mapping a set of performance measures onto a single scale where the standard deviation is one and the mean value is zero (see Hinton 2004; Miles and Shevlin 2006 for detailed accounts of the underlying mathematics).

This again raises important questions such as "Is the process of standardisation incorporated and applied by the compilers in preparation of their rankings though?"

5.3.6 Weighting PIs Prior to Aggregating

Anyone who has had the experience of adding oil to petrol to run a two stoke petrol engine knows exactly what ensuring the correct mix means. If the mix is 25:1, then making the mix 15:1 will result in trouble! In this example, there is a theoretical underpinning, beyond my understanding admittedly, that justifies why 25:1 is the correct mix. How does this relate to university rankings?

In the creation of a single ranking index, several measures are added together. But what is the mix or the weighting applied. It is not difficult to realise that given

Table 5.6 Guardian PI measures and relative weightings used creating the 2010 rankings

We have used seven statistical measures to contribute to the ranking of a university or college in each subject, weighted as follows:

- Teaching quality: as rated by final year students on the course (10%)
- Feedback (assessment): as rated by final year students on the course (5%)
- Spending per student (17%)
- Staff/student ratio (17%)
- Value added: comparing students' degree results with their entry qualifications (17%)
- Entry score (17%)

Source: MacLeod and Hiely-Rayner (2009b)

Table 5.7 Maclean's 2008 weighting and measure

Category	Sub-factor	Weighting (%)
Students/classes	Student awards	20
	Student faculty ratio	
Faculty	Faculty awards	18
	Faculty grants	
Resources	Research income	12
	Operating budget	
Student services	Scholarships and bursaries as % of budget	13
	Student services as % of budget	
Library	Expenditure	15
	Requisitions	
	Holdings per student	
Reputation	Survey	22

Source: Dwyer (2008)

so many global university ranking tables, there is no agreement of the relative contribution of the measures. In simple terms, the weightings adopted by compilers are idiosyncratic and devoid of a theoretical underpinning.

Looking more closely at Table 5.6, what the *Guardian 2010* compilers are implying is that qualifications on entry (17%) are more important in the contribution to the overall ranking index than the teaching quality which contributed 10%. On what basis are these 2% based? Who decides that one measure contributes more to the overall measure of the university?

To show that this is not restricted to the UK, consider the way in which Maclean's university ranking operates for students in Canada. Maclean's, unlike compilers in the UK, places universities in one of three categories, recognising the differences in types of institutions, levels of research funding, the diversity of offerings, and the range of graduate and professional programs. The three categories are primarily: Undergraduate universities where few graduate programs are available; Comprehensive category where there is a significant volume of research and there are many graduate programs on offer; and finally those defined as Medical Doctoral Universities where a broad range of Ph.D. programs and research are provided and where there are medical schools, which set them apart in terms of the size of research grants (Table 5.7).

It might be expected that an analysis of the PIs, that the compilers plan to use, might be statistically analysed to tease out the relative importance of the contributing PIs. It would be possible to use Factor Analysis or Logistic Regression to secure a measure of confidence in the relative importance of the PIs to each other and therefore to the final ranking index.

Why is it the case that compilers do not follow this path?

Combining data may appear innocuous but little research has been conducted that allows us to make a balanced judgment as to the balance that ought to be applied. The principle involved in the process of weightings involves assigning to each indicator a weight that reflects the perceived importance and then combining these weights into an overall score.

> Nonetheless, just as democracy, according to Winston Churchill, is the worst form of government, except for all the others, so quality rankings are the worst device for comparing the quality of … colleges and universities, except for all the others. (Webster 1986)

Shapiro (1994), principal and vice-chancellor of McGill University at the time, commented on the shortcomings of *Maclean's* ranking publication and drew attention to graduation rates. A university with high graduation rates could either be a university 'providing effective education and support to excellent students or a university with lax evaluation and standards'. His letter to the editor of *Maclean's* questions the logic of combining indicators to:

> …obtain a global evaluation or ranking is the most difficult for Maclean's to rationalise. The process requires a decision on the weight to be assigned to each parameter in the equation. These weights must arise from value judgements on which there will never be universal agreement. In Maclean's case, these are based on the values of the Maclean's editors. It is quite clear that a different set of values could result in a different global evaluation and ranking…and it is impossible to determine objectively which set of values and weights is to be preferred (Shapiro 1994).

The compilers start with a mass of data and through a series of mathematical and statistical procedures reduce the data to a single column. In the example provided above relating to how standardisation of scores can assist in the process of aggregating two PIs, the implicit assumption was that the two PIs would be aggregated like for like. The assumption had no theoretical foundation why should *PI* A contribute equally with *PI* B to the overall score.

5.3.7 Creating a Single Index Reflecting a University

The primary objective of the university, subject, or research ranking is to end up with a single measure – a metric – that constitutes a measure of quality. The final stage is then to sort the indices into an order from high to low. No account is taken at this final stage of the significance of any differences between the indices that emerge from the processes described above Richardson (2008: 14).

5.3.7.1 Statistical Difference

Gerhard Casper, then president of Stanford University, in a letter of protest to the editor of the *US News and World Report*:

> "…Could you not do away with ranks ordering and overall scores, thus admitting that the difference between #1 and #2 – indeed between #1 and #'10 – may be statistically insignificant." (Casper 1997)

Clarke, citing a more extensive quotation from this letter, raises the question that while the issue has received much debate but acknowledges that little research has been conducted on the implications.

The significance of difference in scores is not easy to judge from a ranking table where small differences in scores can translate to substantial differences in ranking because of heavy clustering around the mean. The *Times Higher World Ranking of Universities* (QS 2009) in the subject cluster Social science finds a difference of just 1.2 points on a 100 point scale between rank 83 and 89. In the overall university rankings, there is just a 1.9 point difference between rank 64 and 73 going down to a slim 1.2 point difference between rank 65 and 72.

5.3.7.2 Volatile Rankings

Confidence with the constancy of university rankings may be challenged by the data that follows. The data represented in Fig. 5.4 is taken from rankings created by the Guardian Newspaper in June 2009. The graph shows on the left hand scale the rank index order for the UK[4] universities based on the order for 2010. Universities are represented by the column that increase from left to right (light shading); superimposed on that graphic is a secondary graph that reports the difference between the 2010 and 2009 ranking position for each university. It is possible to suggest that:

- Small fluctuations in the size and number of dark bars (indicative of changes between the two years) are indicative of ranking consistent across years. Little change occurs and the rank order is resilient.
- Large fluctuations in both the number and size of the dark bars (indicative of change between the two years) are indicative of turbulence (Longden and Yorke 2009) between the years.

What are the implications of such a volatile system? Is it possible for an institution to change its relative position to other institutions from 1 year to the next? Figure 5.4 shows that fluctuations, or turbulence, occurs randomly among institutions from year to year. The graphic shows 2008 university ranking for UK universities in light grey with dark bars superimposed on the base data for 2008, representing the change between 2008 and 2009 ranking data.

[4]In 2009, five higher education institutions refused to release data held by HESA to compilers involved in the creation of university rankings. There is evidence that this is an increasing trend in the UK. In 1999, there was a single institution refusing to release data.

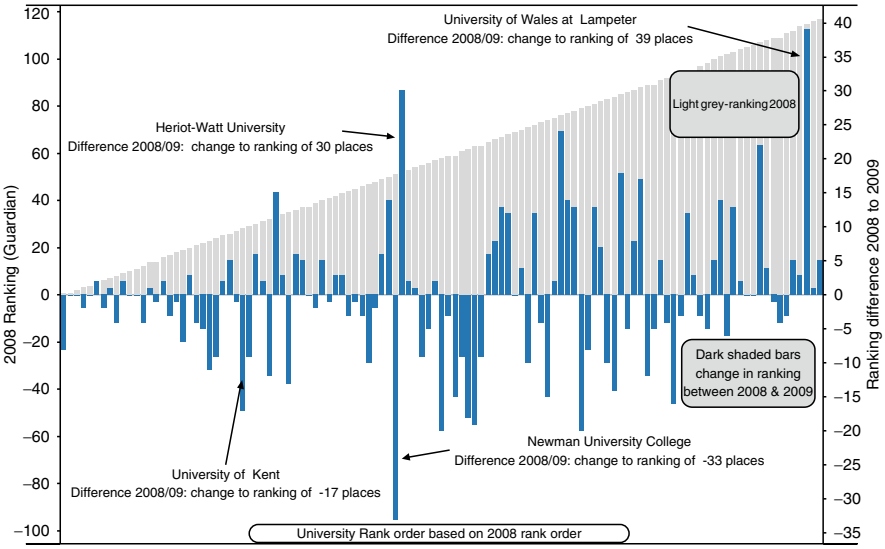

Fig. 5.4 Change in ranking index for UK universities between 2008 and 2009 (Source: MacLeod and Hiely-Rayner, 2009b)

Measurements recorded above the zero base line arise from positive movements in 2009 compared to the base line data for 2008. Measures below the zero base line indicate institutions where the 2009 placement is below that for 2008.

If the system were stable, then there should be few if any dark bars superimposed as the difference would be zero. Interpreting the data in Fig. 5.4 suggests that:

- A large number of dark bars (difference between 2008 and 2009) superimposed over the 2008 ranking, implying that many changes in university rankings occur between the years.
- The length of the dark bars (difference between 2008 and 2009) provides a visual indication of the size of the difference; for example, The University of Wales at Lampeter moved 39 places.
- The dark bars indicate that some of the differences are substantial, both positive and negative, implying that for some universities, the change in ranking is significant.

There are several explanations that can be considered to account for this. It could be related to changes in the methodology between the years in question giving rise to the fluctuation. It could also be related to internal institutional behaviour. The behaviour of Clemson could be considered to cause such a fluctuation, but it might also relate to negative outcome from internal reorganisation.

This raises further important questions that need to be addressed if confidence is to be restored. Should readers be informed about the volatility of rankings and that an institutional ranking may be subject to wide variation between the year the data was collected and the student formally engaging with the institution?

When data prepared by QS for The *Times Higher World Ranking of Universities* (QS 2009) is subject to the same analysis the results for 2007 and 2008 are shown in

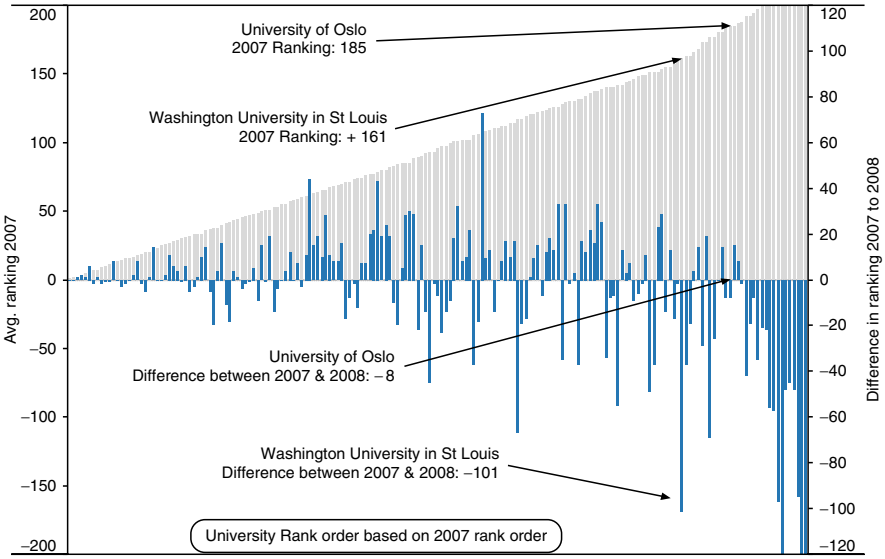

Fig. 5.5 World ranking volatility between 2007 and 2008

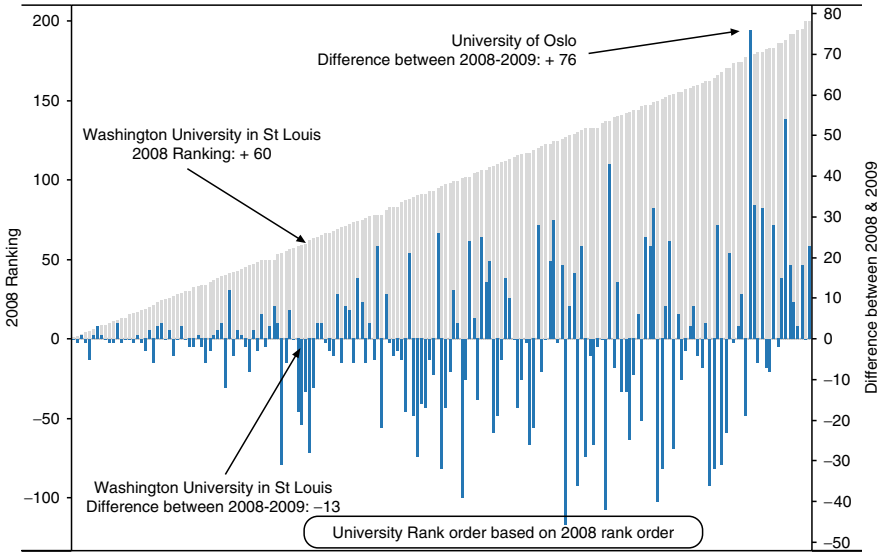

Fig. 5.6 World ranking volatility between 2008 and 2009

Fig. 5.5 and for data relating to 2008 and 2009 in Fig. 5.6. The graphics show clearly that considerable turbulence is evident over the two years. The turbulence appears to be greater in Fig. 5.5 towards the lower rank in the order of universities, although both graphics suggest extensive and substantial variation in differences between two years of data. For example, Washington University in St Louis from 2007 to 2008 moved

down 101 places but moved up 60 places by the time 2009 data was published. A similar pattern can be detected for the University of Oslo, which dropped 8 places between 2007 and 2008 but moved up 76 places by the end of 2009.

The three graphics (Figs. 5.5–5.7) provides evidence that the turbulence is not limited to teaching or research but is evident in both forms of university activity.

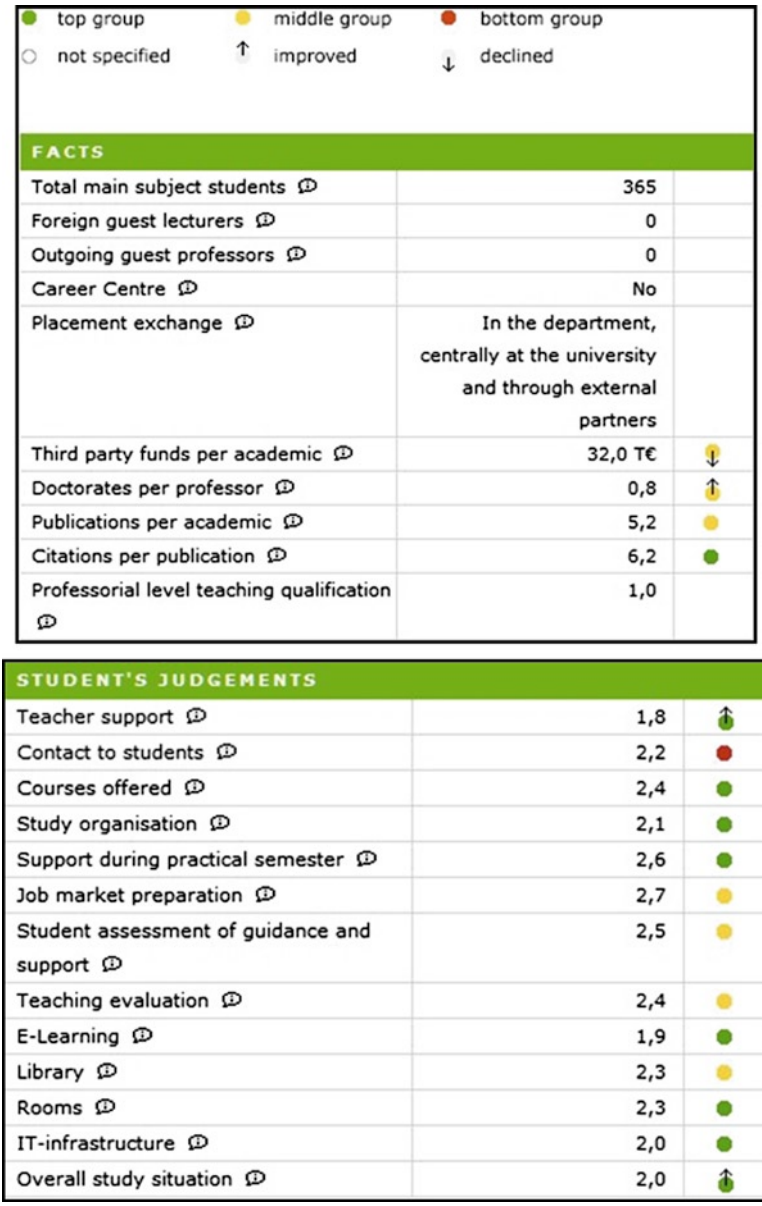

Fig. 5.7 Typical CHE university ranking data for subject 'X' at university 'Y' illustrating the 'traffic light' presentation of ranking data

5.3.7.3 Reality Check

Anecdotal conversations are frequently cited where compilers when asked what they would do, if the final ranking indices were unexpected and elevated an institution substantially beyond the anticipated or expected position, suggest that they would adjust the algorithm. It is often referred to as the reality check. It raises the question about '*whose reality is being used as the benchmark for checking*?' It may even suggest that the algorithm is derived from the expected ranking rather than the other way round!

It also raises the ethical question for the compilers when an unexpected index occurs, 'what should they do next!'

5.4 Positive Developments

The focus in this chapter has been on methodology drawing on examples drawn mainly from rankings published in the UK, USA and Canada. Increasingly, most publishers are making greater use of the potential that publishing on the web can offer. A recent development made possible through the web is the development of an interactive approach, leaving it to the reader to select key indicators in the creation of an overall score. This approach does not vindicate the criticisms discussed in this chapter, but quite the reverse, because it passes responsibility for measuring 'quality' from the publisher to the potential student.

An innovative approach has been developed at the Centre for Higher Education Development (2009) in Germany, designed to address the needs of providing quality information to prospective first-year students and the need to identify research performance quality.

The methodology used for University Ranking (CHE-Hochschul Ranking) relies on data relating to the departmental/subject level in contrast with the usual interest of ranking at the institutional level. By making this decision, CHE-Hochschul Ranking acknowledges that many weaker institutions have national or world class departments that would otherwise be overlooked. It also rejects the concept of 'best HEI'.

At the heart of the methodology is the idea that universities and colleges have individual strengths and weaknesses and that there are no empirical or theoretical bases on which to give weighting to individual factors. It argues that, as the target group is first year students where they are heterogeneous in their preferences, it would be counterproductive to use fixed predetermined weightings.

Instead the HEI is 'viewed' from several different perspectives – professors, managers, students. Each allows for contrast to be made between subjective assessment and objective indicator. Institutions are not given an individual ranking position but assigned to a ranking group of top, middle and end group – which gives the appearance of a traffic light presentation (see Fig. 5.7). A comparable approach has been adopted in the food industry where 'traffic light' graphics are used in food packaging to inform the consumers on food quality.

The ranking, therefore, never tells the user who is the best but maybe who is performing better than average on indicators considered relevant to the user.

The CHE Research Ranking (CHE-Forschungs Ranking) currently covers 16 subjects from natural sciences, humanities and social sciences. It does not define ranking positions but determines the top groups for individual indicators. It is determined based on the following factors:

- Level of third-party funding spent on individual subjects
- Number of doctorates
- Publications and citations
- Patent registrations or inventions

Interviews with professors provide additional information that is given on the reputation of universities with respect to the subjects analysed. However, this information is not used to determine the top groups.

5.5 Demystification and Confidence

The challenge set for this chapter was to explore the methodologies used by the compilers of university, college, and research rankings and to test out to what extent we should have confidence in them.

The view taken early in the chapter was that a negative critique of the methodology does not imply an intention to mislead the reader of such rankings. It was suggested that the ranking methodology is complex and occasionally, compilers are reluctant to de-mystify steps used – for commercial sensitive reasons – and thus, we the users of rankings have to rely on the validity, reliability and consistency of the output from the methodology adopted and applied by the compilers and publishers. Leach (2004), from the perspective of the compilers, comments on the limitations of the university rankings.

> University table or more specifically the rankings we employ, generate a fair amount of anger in the academic community. Institutions are often annoyed at the methodology and the data we choose, and at the sheer gall of marking them against each other in the first place. But we believe that, on balance, tables like these are important. (Leach 2004)

It is clear from the quotation that Leach feels that '… on balance…', there is more to be gained from the tables than lost and as the impact of debt increases, it is important that students '…know what they are getting for their cash'. It is my proposition that the tables really do not provide the answer that they may be searching for.

The *black box* has been opened, the compilers have made available insights into the processes they perform to create the index, yet most of us are unwilling – not unable – to engage in a critical discourse with the compilers to challenge them to provide a justification for each step in the process and to provide a philosophically sound rational justification that allows them to use a single metric to define a university.

The final most critical question remains 'How can a university be reduced to a single metric which is meaningful?'

It remains the single most disconcerting aspect of the whole process of creating a ranking, one that defies logic and one that is so patently wrong. A university is a complex, dynamic organisation constantly changing, year on year with respect to the faculty providing the teaching, to the form and nature of the curriculum offered, to the resources provided. To capture all that complexity in a single measure makes little sense.

Add to this the fundamental methodological criticism described in the paragraphs above where at each stage in the process profound criticism have been advanced at the limited theoretical framework informing assumptions adopted by compilers.

- From the selection of specific events over other specific events
- Their conversion into numerical values
- The adaptation of these numerical values on to scales
- The aggregation of these scaled indices to create a single measure
- A theoretical belief that the measure is capable of defining the quality of a university, a teaching subject, a department, or a research group

Universities are complicit in the process and fail to critically stand out for a more robust and honest attempt at providing information to prospective students rather than play '*our ranking is better than your ranking*'.

A critic of *US News* and *World Report*, Thompson (2000) claims that there is sound evidence that universities and colleges alter their policies for the sake of the rankings – the Heisenberg effect, thus changing the very thing being measured – and giving rise to the danger of mission drift, valuing aspects of university and college life that are exposed to the measurements and thus devaluing those aspects less open to an objective measurement.

Thompson maintains that rankings are:

> … opaque enough that no one outside the magazine can figure out exactly how they work, yet clear enough to imply legitimacy.

A view that accurately summarises the position in 2010.

References

Altbach, P. (2006). The dilemmas of ranking. *International Higher Education*. Retrieved 25 June 2009, from http://www.bc.edu/bc_org/avp/soe/cihe/newsletter/Number42/p2_Altbach.htm

Baty, P. (2009). The world in motion. *Times Higher Education*. December 17, 2009.

Brown, R. (2006). League tables – do we have to live with them? *Perspectives: Policy and practice in higher education, 10*(2), 33–38.

Casper, G. (1997). Private letter from Gerhard Casper, president of Stanford University, to James Fallows, editor of U.S. News & World Report, Personal communication received by J. Fallows, on September 23, 1996.

Cave, M., Hanney, S., Henkel, M., & Kogan, M. (1997). *The use of performance indicators in higher education: the challenge of the quality movement*. London: Jessica Kingsley.

Centre for Higher Education Development. (2009). Retrieved 25 July 2009, from http://www.che-ranking.de/cms/?getObject=613&getLang=en

Clarke, M. (2002). Quantifying quality: What can the U.S. News and World Report ranking tell us about the quality of higher education? *Education Policy Analysis Archives*. Retrieved 20 June 2009, from http://epaa.asu.edu/epaa/v10n16

Clarke, M. (2004). Weighing things up: A closer look at US News & World Report ranking formulas. *College and University Journal, 79*(3), 3–9.

Dwyer, M. (2008). Our 18th annual rankings. Retrieved 4 Dec 2009, from http://oncampus.macleans.ca/education/2008/12/19/our-18th-annual-rankings/

Ehrenberg, R. G. (2003). Reaching for the brass ring: The U.S. News & World Report rankings and competition. *The Review of Higher Education, 26*(2), 145–162.

Graham, A., & Thompson, N. (2001). Broken ranks. Article in *Washington Monthly*.

Harvey, L. (2008). Rankings of higher education Institutions: A critical review. *Quality in Higher Education, 14*(3), 187–208.

HESA. (2009). Student definitions 2007/08. HESA Student Record Retrieved 20 June 2009, from http://www.hesa.ac.uk/index.php/component/option,com_datatables/task,show_file/defs,1/Itemid,121/catdex,3/disp,/dld,institution0708.xls/yrStr,2007+to+2008/dfile,studefs0708.htm/area,institution/mx,0/

Hinton, P. (2004). *Statistics explained*. Hove, East Sussex, UK: Routledge.

Leach, J. (2004). How not to use the tables. Article in *The Guardian*.

Lederman, D. (2009a). Manipulating,‘ Er, Influencing ’*U.S. News*. *Inside Higher Ed*. Retrieved 22 June 2009, from http://www.insidehighered.com/news/2009/06/03/rankings

Lederman, D. (2009b). More rankings riggings. *Inside Higher Ed*. Retrieved 22 June 2009, from http://www.insidehighered.com/news/2009/06/08/usc

Longden, B. (2008). Performance Indicators. In G. McCulloch & D. Crook (Eds.), *The Routledge International Encyclopaedia of Education* (pp. 760).

Longden, B., & Yorke, M. (2009). Institutional rankings, marketing, and the needs of intending students. In B. M. Kehm & B. Stensaker (Eds.), *University rankings, diversity, and the new landscape of higher education*. Rotterdam: Sense Publications.

MacLeod, D., & Hiely-Rayner, M. (2007). University tables 2008: Methodology. *Guardian*. Retrieved 1 June 2009, from http://education.guardian.co.uk/universityguide2008/story/0,,2067150,00.html

MacLeod, D., & Hiely-Rayner, M. (2008). University tables 2009: Methodology. *Guardian*. Retrieved 1 June 2009, from http://education.guardian.co.uk/universityguide2009/story/0,,2067150,00.html

Macleod, D., & Hiely-Rayner, M. (2009a). University guide 2009: The calculations behind the tables. *Guardian*. Retrieved 2 June 2009, from http://education.guardian.co.uk/university2009/story/0,,2276943,00.html

MacLeod, D., & Hiely-Rayner, M. (2009b). University guide 2010: University league tables. *Guardian*. Retrieved 20 June 2009, from http://www.guardian.co.uk/education/table/2009/may/12/university-league-table

Miles, J., & Shevlin, M. (2006). *Applying regression & correlation: A guide for students and researchers*. London: Sage Publications.

Morris, H. (2005). A rank influence. Article in *The Guardian*.

Morse, R. (2009). Undergraduate ranking criteria and weights. Retrieved 29 June 2009, from http://www.usnews.com/articles/education/best-colleges/2008/08/21/undergraduate-ranking-criteria-and-weights.html?PageNr=3

Morse, R., & Flanigan, S. (2006). America’s best colleges 2007. Retrieved 8 May 2007, from http://www.usnews.com/usnews/edu/college/rankings/about/weight_brief.php

Morse, R., & Flanigan, S. (2009). How we calculate the rankings. Retrieved 22 June 2009, from http://www.usnews.com/articles/education/best-colleges/2008/08/21/undergraduate-ranking-criteria-and-weights.html?loomia_ow=t0:s0:a41:g2:r5:c0.212928:b20174593:z0&s_cid=loomia:about-the-rankingsmethodology

Newman, M. (2008). Students urged to inflate national survey marks to improve job options. *The Times Higher*. Retrieved 25 June 2009, from http://www.timeshighereducation.co.uk/story.asp?sectioncode=26&storycode=401883

O'Leary, J., Hindmarsh, A., & Kingston, B. (2006). *The good university guide 2007*. London: The Times.
QS. (2009). QS.com Asian university rankings 2009. From http://www.qsnetwork.com/
Richardson, J. (2008). *Counting what is measured or measuring what counts? – League tables and their impact on higher education institutions in England*. Bristol: Higher Education Funding Council for England: Appendix B.
Richmond, W. K. (1968). *Readings in education*. London: Methuen & Co Ltd.
Ryan, A. (2009). The trouble with tables. Article in *The Times Higher.*
Shapiro, B. (1994). Personal communication from Bernard Shapiro to Ann Dowsett, Personal communication received by A. Dowsett, on 20th September 1994.
Shea, C. (2009). Shocked, shocked: More evidence of rankings shenanigans. *Boston Globe*. Retrieved 4 Dec 2009, from http://www.boston.com/bostonglobe/ideas/brainiac/2009/06/shocked_shocked.html
Stella, A., & Woodhouse, D. (2006). Ranking of higher education institutions. J. Baird: AUQA Occasional Publications Series no. 6. Melbourne: AUQA from http://www.auqa.edu.au/files/publications/ranking_of_higher_education_institutions_final.pdf
Thompson, N. (2000). Playing with numbers: How U.S. News mismeasures higher education and what we can do about it *Washington Monthly*.
Times online. (2009). Good university guide 2010: How the tables work. *Times Online*. Retrieved 20 June 2009, from http://extras.timesonline.co.uk/tol_gug/gooduniversityguide.php
Usher, A., & Savino, M. (2007). A global survey of rankings and league tables. In Institute for Higher Education Policy (Ed.), *College and university ranking system*. Washington, DC: Institute for Higher Education Policy.
Watson, D. (2008). Universities behaving badly. *Higher Education Review, 40*(3), 3–14.
Webster, D. S. (1986). *Academic quality rankings of American colleges and universities*. Springfield, IL: Charles C. Thomas.
Yorke, M., & Longden, B. (2005). *Significant figures: Performance indicators and 'league tables'*. London: Standing Conference of Principals.

Chapter 6
Measuring Faculty Productivity

Karen L. Webber

As with other aspects of higher education, we now have rankings of faculty productivity. A better understanding of the breadth of measures used for faculty productivity is discussed in this chapter.

6.1 Introduction

Due to economic cycles with reduced funding, interest in institutional rankings, and prestige seeking, interest in faculty productivity has gained increasing importance in higher education over the past century. Numerous studies have examined factors that contribute to faculty productivity (Baird 1986; Bellas and Toutkoushian 1999; Bland and Ruffin 1992; Blackburn 1985; Chan et al. 2004; Creamer 1998; Creswell 1985; Dundar and Lewis 1998; Feldman 1987; Kyvik 1995; Levin and Stephan 1991; Marsh and Hattie 2002; Porter and Umbach 2001; Toutkoushian et al. 2007).

This chapter provides a summary of those factors/indicators used to mveasure faculty productivity in post-secondary institutions and the issues related to their measurement. These issues are varied and complex. Typically, faculty members perform myriad tasks that are sometimes difficult to quantify and vary widely across type of institution and discipline; yet, external constituents as well as internal planning processes require more detailed data on faculty productivity. To accommodate this need for more information, institution officials are devising ways to more accurately collect and utilize faculty productivity data.

Although dips in the economy and consequent decreases in budget allocations for higher education from state and other external sources may contribute to some management plans that support academic capitalism (Slaughter and Rhoades 2004), the requests for continued support from states and legislative officials often prompt increased scrutiny of faculty workload and productivity. Recently, some state officials have focused on faculty commitment to teaching, while others have urged

K.L. Webber
Institute of Higher Education, University of Georgia, Athens, GA, USA
e-mail: kwebber@uga.edu

J.C. Shin et al. (eds.), *University Rankings*, The Changing Academy – The Changing Academic Profession in International Comparative Perspective 3,
DOI 10.1007/978-94-007-1116-7_6,

reliance on extramural funds so as to reduce the reliance on state appropriations (Callan 2007; Delaney and Doyle 2007). The demand for increased student credit hours generation is common, as is the heated exploration for external funds that lessen the burden on state and federal funds.

As with other aspects of higher education, we also have rankings of faculty productivity. We rank academic departments, individuals, and institutions based on faculty scholarly products such as publications, presentations, and patents. A number of measures exist including the faculty productivity index by Creswell (1985), the index of research productivity (IP) by Ramsden (1994), academic research productivity by Jordan, Meador, and Walters (1989), and the more recent h-index by Hirsch (2005). Faculty effort in research and in some cases teaching is an important part of some institution ranking schemes, including the Academic Ranking of World Universities, the QS/Times Rankings (Quacquarelli Symonds 2009), The National Research Council's Ranking of Graduate Programs, and *US News and World Report* College Rankings. Even though some rankings for faculty productivity are done at the individual level (for example, Cronin and Meho 2006), most are done at the program or institutional level. *Academic Analytics* is a privately owned company that seeks to provide analytical data on faculty productivity to universities and has published faculty department rankings (*Chronicle of Higher Education* article, January 12, 2007) based on productivity measures as defined by their company. Measures such as book and journal publications, journal article citations, federally funded research grants, and honors and awards are used to calculate a Faculty Scholarly Productivity Index (FSP Index). According to officials at *Academic Analytics* (*personal communication on January 13, 2010*), data is primarily used by their clients for strategic planning purposes. The database can be sorted and organized in a variety of ways to allow the client to determine the department or institution's rank based on different criteria, and thus have the ability to accommodate for differing emphases in the program or institution.

6.2 The Challenges of Measuring Faculty Productivity

There are a number of consistent themes and measures among the inquiries on faculty productivity, but differences do exist. Although most discussions of faculty productivity acknowledge the nuances and need for more qualitative measures, the challenges in the use of qualitative measures have prompted most institutions to rely on quantitative measures. Issues that affect all faculty members include tasks related to teaching (face-to-face or distance instruction), advising, faculty governance, and other committee work. Most of these tasks can be quantified in terms of number of students taught or committees served; yet, often the total time spent with a student or the amount of effort devoted to a new instructional technique or course are often hard to quantify and do not address the quality of effort at all. Similarly, the total effort devoted to the initial laboratory set up, introduction of new graduate lab assistants, becoming familiar with lab protocols, and learning to work as a

collaborative team member may not be accounted for fully in the simple measures of workload and thus may get marginalized in productivity counts.

The relative degree to which faculty members are evaluated for teaching, research, and public service depends on the mission of the institution as well as the context and discipline. For example, the proportion of time spent on teaching for a community college professor is vastly different than that of the research university peer, and thus workload products are also different. Even within one institution, the proportion of time spent on teaching versus other scholarly activity will vary; typical faculty members in business receive far less extramural funding and spend less time on traditional research compared to faculty members in biology.

6.3 Conceptual Models of Faculty Productivity

There are several conceptual models that can be employed to examine faculty productivity, many of which are discussed in greater detail in other chapters of this book. These models include a more thorough discussion on differences in productivity by gender, race, time status, and type of institution, the tensions of teaching and research, and the efficacy and future of tenure. Although discussions of faculty workload are lively at all US institutions, the discussions on faculty productivity related to research and service emanate primarily from the four-year sector, thus the majority of issues discussed in this chapter relate to comprehensive measures of faculty productivity at four-year institutions.

Some scholars have focused more narrowly on one or two facets of faculty productivity. For example, Bailey (1999) and Blackburn et al. (1991) examined the level of motivation and self-efficacy related to teaching, research, and service. Golden and Carstensen (1992) studied the effect of institution control (private vs. public) and size of department on per capita publications, while Crosta and Packman (2005) examined faculty productivity through supervision of doctoral students. Several authors, including Dusansky and Vernon (1998), Dwyer (1994) and Hagerman and Hagerman (1989) focused on a count of publications, and Middaugh's initial work (2001) primarily examined productivity through student credit hours generated.

Others have taken a broader look. Dundar and Lewis (1998) reported that faculty productivity is based on a series of individual and institutional attributes. Massy and Wilger (1995) defined productivity as the ratio of outputs to inputs, or benefits to costs. Fox (1985) grouped factors according to three main clusters: individual ascriptive (age and gender); Individual achievement (including rank, years in higher education, quality of graduate training, hours spent on research each week, and extramural funds received); and institutional characteristics. And Harris (1990) purported that research performance can be evaluated through four facets: impact, quality, performance, and quantity. He measured impact and quantity by counting the number of citations made and referenced by other scholars as well as the

importance and quality through judgments such as peer review. While each of these researchers takes a slightly different perspective, they all acknowledge the need to include some factors related to the individual as well as some factors related to the environment.

Also looking broadly, Porter and Umbach (2001) believe faculty productivity can be grouped into five areas: individual demographics; teaching workload; career status; personal career preferences; and dimensions of human capital including knowledge, skill values, education, and training. Porter and Umbach as well as many others (Dundar and Lewis 1998; Fairweather and Beach 2002; Feldman 1987; Golden and Carstensen 1992; Long 1990) acknowledge the interplay of factors such as extramural funds received, size of academic department, number of "star" faculty, and mentor experiences early in one's career.

Somewhat similarly, Creswell (1986) grouped research performance in three categories of individual, department or program, and institutional measures. Bland and colleagues (Bland and Ruffin 1992; Bland et al. 2002, 2006) proposed that faculty productivity is a complex blend of individual, institutional, and leadership factors. Bland and Ruffin (1992) identified 12 common characteristics of a productive research environment: 1. clear goals that serve a coordinating function; 2. research emphasis; 3. distinctive culture; 4. positive group climate; 5. assertive participative governance; 6. decentralized organization; 7. frequent communication; 8. accessible resources, particularly human; 9. sufficient size, age, and diversity of the research group; 10. appropriate rewards; 11. concentration on recruitment and selection; and 12. leadership with research expertise and skill in both initiating appropriate organizational structure and using participatory management practices (p. 385). Like most other aspects of the institution, it seems quite possible that faculty involvement spans a broad spectrum of activities and events. When the activities, time invested, and resultant rewards are varied and differ depending on the individual and institution, it becomes clear how the measure of faculty productivity can become complex.

Within the broad discussion of workload and productivity, some scholars have examined the potential synergy between teaching and research (Braxton 1996; Braxton and Hargens 1996; Creswell 1986; Fairweather 2002; Feldman 1987; Fox 1992; Harry and Goldner 1972; Linsky and Straus 1975; Ramsden and Moses 1992) or teaching, research, and service (Blackburn et al. 1991). Although some propose that a positive synergy can or should come from interaction of the two activities (Marsh and Hattie 2002), or a possible curvilinear relationship (Harry and Goldner 1972; Levin and Stephan 1991; Linsky and Straus 1975), three major meta-analyses found only a small (nonsignificant) overall relationship between the two activities (Hattie and Marsh 1996; Feldman 1987; Marsh and Hattie 2002). Yet, while Marsh and Hattie admit these two activities are "at best loosely coupled" (p. 606), Braxton and Hargens (1996) suggest that the teaching–research relationship differs by department and is more strongly related in some disciplines. To support this argument, they refer to Feldman's (1987) findings that the average correlation between teaching and research was 0.22 for humanities, 0.20 for social sciences, and 0.05 for natural sciences.

Time and again, the context of the discipline surfaces as an important variable. Several scholars including Adam and Roberts (1993), Baird (1986) and Becher (1994) discuss the importance of examining productivity by discipline. Still, today, the use of Biglan's (1973) grouping scheme is common. With clusters on three dimensions: hard versus soft; life versus nonlife; and pure versus applied, Biglan found differences in the level of social connection within the academic department, collaboration with peers, commitment to teaching, and allocation of time in academic tasks. Similarly, Fairweather and Beach (2002) point out "the futility of portraying an 'average' research university because of the variances across disciplines" (p. 101). Bland and colleagues also speak about the importance of the culture and in particular, academic discipline. For example, Bland and Ruffin (1992) report that higher faculty productivity is seen in those departments that provide mentorship by established faculty. Disciplinary norms are also important; it is not uncommon for a professor in a biology or chemistry department to coauthor 10 or more articles per year, while a highly productive professor in education may author only two or three. The collaborative team culture that exists in science lab settings provides a larger workforce and expectations that all will contribute to each phase in the research process.

Acknowledging the complexities of differential expectations, tasks, and cultures, a multifaceted look at factors affecting faculty productivity across a broad spectrum is most appropriate. Figure 6.1 delineates the measures used to examine faculty productivity across the dimensions of research, instruction, and service. However, it is important to acknowledge the complexities that arise across different dimensions, and these measures of faculty productivity must be considered within the context of the individual, the institution, and the environment.

6.3.1 Indicators of Faculty Productivity Related to Research

The broadest set of productivity measures are related to a faculty member's research, perhaps in part because many of these measures are easier to quantify. Traditionally, faculty productivity is discussed through the enumeration of research productivity, and research productivity in turn has most often been defined as number of publications produced in a short period of time (such as one or two years) or a lifetime career. For example, Feldman's (1987) review found that 21 of 29 studies used the number of publications as the measure of productivity. Creswell (1986) reminds us that the "publish or perish" mentality prompts us to count the number of publications but that publications are only one measure of research. Similarly, we might then consider the fact that publications are only one form of research productivity.

Although research productivity is the most widely discussed facet of faculty productivity, the alignment of research investment and productivity with institutional mission is equally important. For example, faculty members at two-year colleges are not generally expected to produce publications, nor are those at

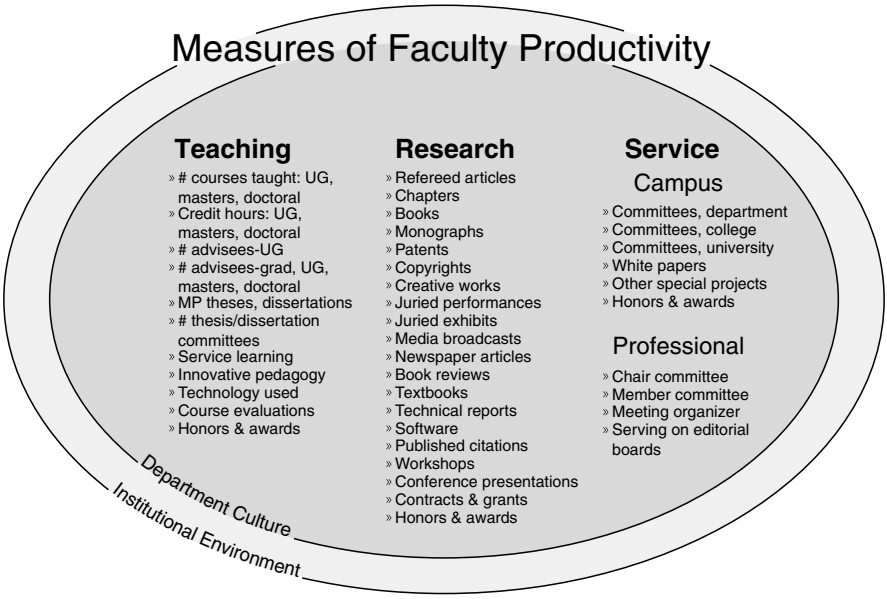

Fig. 6.1 Measures of faculty productivity, shown within the context of the discipline and institutional environment

four-year comprehensive colleges expected to publish at the same rate as those at research universities. Especially at research universities, publications and extramurally-funded grants are central to the institutional image and pocketbook, and thus strongly affect individual promotion and tenure, work satisfaction, and self-esteem.

Indeed, there is much written on quantitative measures of research productivity. As listed in Fig. 6.1, factors may include a straight count on the number of publications (including refereed and/or non refereed books, book chapters, articles, research monographs, commentaries, debates, media broadcasts, book or other reviews), unpublished papers or monographs, number of research grant proposals submitted and/or received, creative works (including plays written, produced, and/or performed, juried or nonjuried exhibits, and literature readings), patents applied and/or granted, computer software developed for private or public markets, workshops developed, and conference presentations. Some institution officials or scholarly studies may also include number of citations, such as those tracked by the Institute of Scientific Information (ISI) index or Web of Science, however, such citation counts may be problematic for individuals who change their name or institutional affiliation, or have coauthored publications. In addition, most of the electronic citation indices do not count all forms of publication (such as book chapters and many forms of humanities products such as exhibits, readings, and artistic performances); thus, one must be cautious if using a citation index as the indicator of research productivity.

Journal and other print publications are often the most frequent indicator of productivity. Publications, however, differ in their value by discipline, and they are rarely "counted" equally. Many questions arise, including the unit of publication (for example, are books counted equal to journal articles), are refereed articles weighted higher than non-refereed articles, and are there different levels of refereed journals? How should technical reports and manuals be counted? Do book reviews count? How should joint-authored publications be included? And if so, with what weight should conference presentations, performances, exhibitions, and letters to the editor be given? In any study, these questions must be acknowledged, determined which documents will be included, and if each will count equally. Noting caution in the count of publications, Braxton and Bayer (1986) recommend the use of a weighting system tailored to the specific discipline and based on criteria determined by individuals within the discipline.

Acknowledging the challenges of the measures to be used, Ramsden and Moses (1992) developed two indicators of individual research performance. They defined an index of research productivity (IP) as the five-year sum of single or multi-authored books, number of papers published in refereed journals, number of edited books, and number of chapters in refereed books. Their index of research activity (IA) was calculated using responses to questions on whether the individual faculty member had or had not participated in the following activities during the past two years:

- Received an external research grant
- Received an internal research grant
- Supervised one or more honors or masters students
- Supervised one or more PhD students
- Had informal discussions with departmental colleagues about research interests
- Participated in one or more joint research projects with colleagues
- Served as editor or on editorial board of an academic journal
- Reviewed one or more proposals for a funding agency
- Refereed one or more articles for a journal
- Delivered one or more conference papers
- Maintained professional contact with colleagues overseas

Although Ramsden (1994) points out that this collection of factors and indices does not include certain aspects of research performance such as quality, importance, commitment, nor satisfaction, they do allow for a more comprehensive look at quantitative factors related to productivity and allow for the development of a model of research productivity.

As mentioned earlier in this chapter, the issue of academic field or discipline is of particular importance when measuring an individual's research productivity. For example, social scientists typically produce 2–3 refereed publications in a year, while life and physical science and related scientists typically produce 3–4 times that number. The differences may have to do with the types of publications (technical reports and essays compared to longer full empirical studies), but may also

involve interacting factors including coauthorship, number and productivity of other faculty in the department, and factors related to the institutional culture.

For many, the number of publications and/or citations is frequently used as one, if not the central, indicator of productivity. There are several derivations on a Publications Impact Factor, and Garfield and Sher are credited with introducing the notions of a Journal Impact Factor in the 1960s (Garfield 2006). Among the iterations in studying the impact of publications, some note that the discipline and quality of journals within disciplines vary; thus, straight counts may not be most accurate. For example, Seglen (1997) argues that journal impact factors differ by research field and do not account for variations in scientific quality of some articles, and Wardle's (1995) analyses found that a disproportionate number of citations were from North American authors. A more recent measure of individual faculty productivity has been developed by Hirsch (2005, 2007). He asserts that his H-index is better than a simple count of published articles or citations because it is a broader measure used to indicate one's total body of scholarly work. Although the H-index is just beginning to be tested across disciplines, it receives positive comments (Bornman and Daniel 2009). The debate on journal impact and citation counts is likely to continue in the future.

Along with, and closely correlated with publications, extramural funding is often included as a measure of research productivity. Extramural funds, patents, and software licenses are important outcomes that indicate a level of productivity and ultimately lead to a level of prestige for the institution. In some instances, internal grant funds may be counted along with externally generated funds (Ramsden and Moses 1992).

Some scholars are quick to point out that extramural grant funds are often strongly correlated with other research measures (Blackburn and Lawrence 1995; Fairweather 2002; Porter and Umbach 2001). Those faculty members who receive grant funding may have greater opportunity to work with graduate assistants, receive release from some teaching responsibilities, and subsequently spend more time on research. The assistance of graduate students and release from teaching most often enables the faculty member to focus on research and produce more articles or other research productivity measures.

6.3.2 Indicators of Faculty Productivity Related to Instruction

Perhaps due to the challenges faced in operationalizing the full spectrum of tasks related to instruction, the measures of faculty productivity for instruction are fewer that those for research. Only those who have taught a college course understand the extent of time needed to prepare well before walking into class. In addition to routine preparation for class that includes reading primary and secondary documents, staying abreast of current trends, making in- and out-of-class assignments, and grading, a faculty member typically advises undergraduate and/or graduate students, communicates with distance students via email or electronic class

discussions, guides students in laboratory or art work, and participates and/or chairs theses and dissertations. It is quite possible that a faculty member may spend three to five times the number of hours preparing for class and advising students for every hour spent in the classroom. Often these other hours of preparation and advisement get lost in the discussions of faculty effort.

Numerous scholars have examined the relationship between instruction and research (Braxton 1996; Crosta and Packman 2005; Blackburn and Lawrence 1995; Creswell 1985; Fairweather 1996; Fox 1992; Finkelstein 1984; McDaniel and Feldhusen 1970; Linsky and Straus 1975). According to Fairweather (2002), measures of instructional productivity serve to meet internal and external accountability needs, and can serve as a proxy of student learning. However, Menges (1981) and Bock (1997) argue that counts of student credit hours should not be a measure of learning. For example, in a 300-person lecture course, it is difficult to elicit discussion, and nearly impossible to allow each student to demonstrate their ability to synthesize information on essay exams. If the highest number of credit hours generated is the goal, it is possible that actual learning will not occur to the fullest extent, but such students will gather minimal information passively. Instead, Menges and Bock believe institutional officials should seek to better quantify and make faculty accountable for meeting learning goals.

Some scholars argue that the tasks of instruction and research complement each other and engender synergy; others find no effect, and yet others report that substantial effort devoted to one diminishes effort and outcomes related to the other. Blackburn (1974) argues that unsatisfactory classroom performance may result from neglecting teaching responsibilities in order to pursue research activities. Marsh (1987) posited that the ability to be an effective teacher and productive researcher are positively related, but Marsh and Hattie's meta-analysis did not support that claim. Ramsden (1994) claims that teaching and research are not mutually related to a single domain, and others (Marsh and Hattie 2002) purport that the two activities require different preparation, involve different personality characteristics, are funded from different sources, and are fundamentally different tasks. Similarly, Barnett (1992) claimed that teaching and research are "inescapably incompatible."

Some scholars have charted the relationship between credit hours produced and number of publications. Hattie and Marsh's meta-analysis of 58 articles found an overall correlation of .06 and say that "at best, teaching and research are loosely coupled" (p. 606). Feldman's (1987) meta-analysis of 29 studies found a significant (although, perhaps not of practical significance) correlation of $r = .12$ between research productivity and student assessment of teaching effectiveness. Braxton (1996) furthered Feldman's analysis by grouping the 29 previous studies (plus one more) into three categories of no relationship, a complimentary relationship, or a conflicting relationship. With strength of the correlation between the studies serving as the criteria, Braxton found 11 of the 30 studies with no substantial correlation between teaching and research, 18 of 30 with a substantial correlation between the two, and only one study with a negative relationship between teaching and research. Clearly, this topic has captured much attention, but has not yet resulted in a definitive answer. Knowing that new technologies

and pedagogical techniques must be addressed speaks to the complexities of the debate on teaching-research compatibility, and it is likely that studies on this topic will continue for some time.

Despite the complexities, there are a few discrete measures of faculty productivity related to instruction, shown in Fig. 6.1. These include the number of courses taught, number of students in each course, number of hours per week devoted to instructional duties, and/or student credit hours generated. Due to the difference in level of preparation and course activities, an attempt is most often made to differentiate graduate from undergraduate instruction. Because of the disproportionate amount of time spent in one-on-one instruction, independent study with one or a small number of students may also be counted, as well as an indication if the independent study is at the graduate or undergraduate level.

Student credit hours generated is one of, if not, the most frequently cited quantitative measures of instructional productivity. Breakdowns may include credit hours by faculty rank and/or time status, and may differentiate credit hours generated by type of instruction such as traditional lecture, laboratory, clinical, independent or directed study, and graduate level instruction for research. An example of such a credit hour report is shown in Table 6.1. Rankings of credit hour production within a department or institution, or as compared to a regional or national norm may be developed or required by constituent groups, as they seek evidence of high workload.

In addition to credit hours generated, the number of advisees, differentiated by level of student (undergraduate, masters, doctoral), as well as thesis and dissertation committee work may also be included as teaching productivity measures. In some institutions, serving as the major professor on theses or dissertations is weighted more heavily, and in some institutions an intense level of work with advisees may be substituted for an instructional course in one's full load count.

As technologies advance and pedagogical thinking shifts, some faculty members are also including measures of technology use, innovative instructional strategies, and efforts that assist in global knowledge in their faculty productivity efforts. Because of the breadth and depth of these items, measures exist as a simple count (count of times included, number of various strategies used, and/or hours invested in one or more of the activities). Examples may include the number of courses that use problem-based learning, number of courses that have a service learning component, the number of contact hours devoted to service learning, or number of

Table 6.1 Sample student credit hour report fall 2006

Course level	Tenure/ tenure track	Full-time Other	Part-time	Graduate assistants	Others	Total credit hours
Lower div	23,318	24,158	12,398	34,988	451	95,313
Upper div	77,589	24,106	17,558	11,621	621	131,495
Total UG	100,907	48,264	29,956	56,609	1,072	226,808
First prof	16,590	4,787	1,104	0	63	22,544
Grad	32,795	3,242	3,921	18	425	40,401
Total	150,292	56,293	33,981	45,627	1,560	287,753

courses that are have a specific percentage of time devoted to international content. Institutions that offer distance courses may also require faculty members to keep track of internet synchronous or asynchronous communications, evidence of course exercises or other activities, such as hours spent in email communication.

Although a point of debate, faculty evaluations by students and/or the department chair may also be used as a measure of instructional productivity. Evaluation forms completed by students not only include feedback on general aspects of the course per se but also perceptions of the instructor as an authority in his/her field, his/her management of the classroom, and his/her ability to communicate the course material effectively. One or both aspects of student evaluations of faculty members may also be included as a measure of faculty instructional productivity. Indeed, there is an intuitive association between faculty teaching and research, but the link is generally not very strong. Linsky and Strauss (1975) report that although there is 'face validity' between instructor knowledge and scholarship, the overall correlation of only .04 between overall teacher rating and total publications score. These authors further state that although there are "meaningful patterns" of association, the absolute value of the correlations (between teaching and research performance) is generally low (p. 97). Similarly, Feldman's (1987) review of nearly 30 studies on the relationship between teaching and research found, though in an overall positive direction, was statistically insignificant.

Typically, measures of instructional productivity are cast in numerics, but increasingly, and perhaps because of the intuitive link between teaching and research not yet confirmed statistically, some faculty and college leaders are seeking better ways to provide qualitative evidence. Teaching portfolios and parts of tenure dossiers attempt to show the multifaceted efforts a faculty member undergoes with student learning and may include details on service learning projects, student comments from problem-based activities, exam essays that identify a match with learning objectives, student comments from end of term evaluations, peer observations, and department chair comments. Collectively, these indicators can paint a picture of the effort a faculty member may put forth in teaching.

6.3.3 Indicators of Faculty Productivity Related to Campus and Community Service

Of the three main areas of productivity discussed in this chapter, measures of campus and community service receive the least attention and vary the widest. As higher education institutions have increased their numbers of administrative professionals, faculty members have moved further away from day-to-day involvement in administrative tasks. Committee assignments are usually the most common form of campus service, while consulting and professional association tasks capture a faculty member's external service time. With a positive attitude, most faculty members take on a service role with some level of positive participation and hope to make a difference, but often the slothful pace of committee work render it to be seen as a necessary evil.

Blackburn and Lawrence (1995) report that no agreement exists on how to assess a faculty member's service performance and there are few scholarly studies on the service components of productivity; however, from what is available, the majority of findings indicate that more time devoted to administrative responsibilities reduces productivity (Meyer 1998; Nettles and Perna 1995) How service activities contribute to promotion and tenure vary widely across institutions, but Blackburn and Lawrence report that personal preference for service and perceived institutional expectations for participation are the most important determinants of participation.

Interestingly, however, examining data from the National Center for Research to Improve Postsecondary Teaching and Learning survey (NCRIPTAL), Blackburn et al. (1991) found that the proportion of variance (in a regression analysis of factors contributing to faculty work) accounted for "significant and larger than anticipated for such undervalued activity" (p. 411). They found that activities such as serving on campus committee, being involved in curriculum revision, and chairing committees are not often described as desired activities; yet, especially at research institutions, a sizable amount of the variance was accounted for by these factors.

Institutional and department missions may have a substantial impact on what activities are considered acceptable. State institutions typically carry a stronger service mission and it is most likely that faculty members at these institutions apportion more of their time and activities to service. The proportion of time (and products counted for productivity) may vary greatly in institutions with cooperative extension offices or other significant community agreements. In addition, each faculty member's support for travel (either through extramural funding or general institutional resources) may greatly alter the number of outcomes in campus and community service.

Acknowledging the great variance, faculty productivity in service can be seen through campus-based activities, interactions within the community, and through professional association activities. On campus, faculty members may serve as a member of a department, division, or college-wide committee. The numeric count of committee memberships may be the measure selected, and most often, chairing a committee is given greater weight. Faculty may also supply white papers or participate in other special projects on campus and again a count of these activities may serve as contribution to service. Many institutions also bestow honors and awards to faculty for community service, similar to awards for high quality teaching and/or research. If a faculty member receives a prestigious honor or award, it may also be counted as an extension within service, teaching, or research (depending on the institutional decision and/or nature of the honor or award).

Faculty member's efforts with professional associations are typically included in their measure of service productivity. Faculty members may chair or participate in committees (again with chair duties weighted more), serve as a meeting organizer, or present papers or workshops to other members of the organization. The presentation of scholarly papers at a conference provides the faculty member with needed feedback on his or her most recent work as well as providing a forum in which to share new information. In addition, faculty members are expected to continue

professional development throughout their career, and that take place through attendance at conferences to learn from colleague presentations as well as longer workshops and training sessions. As with other service activities, each presentation or attendance at a workshop is typically counted numerically and depending on the institution or department, may be weighted more or less. If for example, institution leaders have expressed the goal to increase dialogue and collaboration with international colleagues, presenting scholarly work at an international event may be weighted more heavily than one in the local region.

At times, the distinction between teaching, research, and service may not be clear, especially for some members with broad responsibilities. For faculty members in public service divisions, the number of presentations that are considered for service productivity may be much higher and include a broader variety of activities. Cooperative extension agents, for example, may keep extremely detailed records on the number of phone, email, and face-to-face contacts made within the community. In addition, such agents may be prolific writers for newspaper or newsletter articles, or civic presentations. Some institution officials may count these as service, while others may count them as measures of teaching or general scholarship.

6.4 Strategies for the Collection of Faculty Productivity Data

Due to increasing calls for public accountability, and in attempts to use the data on faculty productivity effectively, campus officials often organize data in a database or repository. Some officials have chosen to develop data collection systems in-house, while others have chosen to purchase a vendor-based product that has a pre-designed structure and perhaps even guidelines that help institution officials implement such a system. These web database applications allow faculty members to maintain their teaching, research, service, experience, professional development, and, in some cases, assessment records. Administrators may use these faculty management systems to support accreditation, build rubrics to assess learning goals and objectives, customize reports (including CVs), create web surveys and evaluations, and benchmark faculty productivity. Two such vendors with products that are gaining momentum in US Higher Education institutions are Digital Measures (www.digitalmeasures.com) and Sedona (www.sedonaweb.com). Good examples of institutionally-developed systems include those at The Ohio State University, University of Central Florida, and The University of Mississippi. As mentioned above, Academic Analytics collects data as well as publishes a Faculty Scholarly Productivity Index (FSPI). More information about Academic Analytics can be found at www.academicanalytics.com.

Regardless of whether an in-house system is developed or a vendor product is purchased, the issue and ranking of faculty based on productivity is not likely to wane in the near future. This requires the accurate collection of data on faculty work, and a campus-wide discussion and standard set of definitions for each task. For example, what is deemed a publication can differ widely and a collective set of

individuals on the campus need to determine if a publication should include all written products, or only those receiving peer review, and/or a specific level of review by certain noted scholars in the field.

6.5 Summary

Interest in faculty productivity continues to be an important topic for internal and external constituents in higher education. The degree to which faculty members are evaluated for teaching, research, and public service depends on the mission of the institution as well as the context and discipline, and even within one institution, the proportion of time spent on teaching versus other scholarly activity will vary by discipline and emphasis on select portions of the institutional mission. The emphasis placed on faculty productivity varies by institution, level of institution, by discipline, and by select demographics characteristics such as race and gender. Our focus on quantitative measures has important implications for the facets of faculty work not easily measured. Differences that may exist by institution type, discipline and demographic characteristics prompt us to consider how the traditional measures may disadvantage some individuals in disciplines that do not follow the traditional models of publications and extramural funding as primary measures for faculty work.

Faculty productivity is often included in the increasingly discussed rankings. In most cases, faculty productivity is part of department and/or institution-level rankings. For example, the Academic Ranking of World Universities and The National Research Council's *Survey of Graduate Programs* use faculty citations as a key data point for their institution and departmental rankings, respectively. Programs and institutions that are ranked higher can attract highly able students, star faculty and a greater chance at securing research patents and extramural funds. Rankings seem to be an inevitable part of today's universities, and faculty productivity will continue to play a critical role.

Although there are several conceptual strategies employed to examine faculty productivity, this chapter serves as a summary to elucidate the measures of faculty productivity and is organized along the dimensions of research, instruction, and service. Because of the challenges found in measuring some aspects of faculty work, quantitative counts of activities are the most common measures. In addition, the largest array of quantifiable measures of workload are related to faculty member research and continue to uphold that current reward structure that finds "published research as the common currency of academic achievement" (Bok in Rau and Baker 1989: 163).

While some scholars focus on one or two subsections, the majority of scholars who examine workload acknowledge the breadth and depth of academic work and thus attempt to include measures across the three dimensions of research, instruction, and service. The measures discussed in this chapter provide an insight but also await refinement from scholars in the field.

References

Academic Rankings of World Universities. (2009). Academic Rankings of World Universities. http://www.arwu.org/. Accessed on 15 Dec 2009.

Adam, B., & Roberts, A. (1993). Differences among the disciplines. In R. Birnbaum (Ed.), *Faculty in governance. New directions for higher education* (Vol. 81). San Francisco: Jossey Bass.

Bailey, J. G. (1999). Academics' motivation and self-efficacy for teaching and research. *Higher Education Research and Development, 18*(3), 343–359.

Baird, L. L. (1986). What characterizes a productive research department? *Research in Higher Education, 25*(3), 211–225.

Barnett, R. (1992). *Improving higher education: Total quality care*. Bristol, PA: Open University Press.

Becher, T. (1994). The significance of disciplinary differences. *Studies in Higher Education, 19*(2), 151.

Bellas, M., & Toutkoushian, R. (1999). Faculty time allocations and research productivity: Gender, race and family effects. *Review of Higher Education, 22*(4), 367–390.

Biglan, A. (1973). Relationships between subject matter characteristics and the structure and output of University departments. *The Journal of Applied Psychology, 57*(3), 204–213.

Blackburn, R. T. (1974). The meaning of work in academia. *New Directions for Institutional Research, 2*, 75–99.

Blackburn, J. (1985). Faculty career development: Theory and practice. In S. Clark & D. Lewis (Eds.), *Faculty vitality and institutional productivity* (pp. 55–85). New York: Columbia Teachers College Press.

Blackburn, R. T., & Lawrence, J. H. (1995). *Work: Motivation, expectation, satisfaction*. Baltimore: The Johns Hopkins University.

Blackburn, R. T., et al. (1991). Faculty at work: Focus on research, scholarship, and service. *Research in Higher Education, 32*(4), 385–413.

Bland, C., & Ruffin, M. (1992). Characteristics of a productive research environment. *Academic Medicine, 67*(6), 385–397.

Bland, C. J., et al. (2002). One school's strategy to assess and improve the vitality of its faculty. *Academic Medicine, 77*, 368–376.

Bland, C. J., et al. (2006). The impact of appointment type on the productivity and commitment of full-time faculty in research and doctoral institutions. *Journal of Higher Education, 77*, 89–123.

Bock, M. T. (1997). *The measure of professorial productivity: Using student learning outcomes criteria*. Paper presented at the Association for the Study of Higher Education, Albuquerque, NM.

Bornman, L., & Daniel, H.-D. (2009). The state of h-index research. *European Molecular Biology Organization Reports, 10*(1), 2–6.

Braxton, J. M. (1996). Faculty teaching and research: Is there a conflict? *New Directions for Institutional Research, 90*(2), 1–96.

Braxton, J., & Bayer, A. (1986). Assessing faculty scholarly performance. In J. W. Creswell (Ed.), *Measuring faculty research performance. New directions for institutional research* (Vol. 50, pp. 25–42). San Francisco: Jossey-Bass.

Braxton, J., & Hargens, L. (1996). Variation among academic disciplines: Analytical frameworks and research. *Higher Education: Handbook of Theory and Research, 11*, 1–46.

Callan, P. (2007). Looking under the hood of higher education. Retrieved 17 Sep 2009 at http://www.insidehighereducation.com/layout/set/print/views/2007/11/02/callan.

Chan, K., Chen, C., & Steiner, T. (2004). Who is publishing? An analysis of finance research productivity in the European region. *Journal of Business Finance and Accounting, 31*(3), 401–437.

Creamer, E. (1998). Assessing the faculty publication productivity: Issues of equity. *ASHE-ERIC Higher Education Reports, 26*(2), 1–91.

Creswell, J. (1985). *Faculty research performance: Lessons from the sciences and the social sciences*. (Vol. 4). Washington, DC: Association for the Study of Higher Education.
Creswell, J. (1986). Concluding thoughts: Observing, promoting, evaluating, and reviewing research performance. *New Directions for Institutional Research, 50*, 87–101.
Cronin, B., & Meho, L. (2006). Using the h-index to rank influential information scientists. *Journal of the American Society for Information Science and Technology, 57*(9), 1275–1278.
Crosta, P. M., & Packman, I. G. (2005). Faculty productivity in supervising doctoral students' dissertations at Cornell University. *Economics of Education Review, 24*(1), 55–65.
Delaney, J., & Doyle, W. (2007). The role of higher education in state budgets. In K. Shaw & D. Heller (Eds.), *State postsecondary education research: New methods to inform policy and practice*. Sterling: Stylus.
Dundar, H., & Lewis, D. (1998). Departmental productivity in American universities: Economies of scale and scope. *Economics of Education Review, 14*(2), 119–144.
Dundar, H., & Lewis, D. R. (1998). Determinants of research productivity in higher education. *Research in Higher Education, 39*(6), 607–631.
Dusansky, R., & Vernon, C. J. (1998). Rankings of US economics departments. *The Journal of Economic Perspectives, 21*(1), 157–170.
Dwyer, P (1994). Gender differences in the scholarly activities of accounting academics: An empirical investigation. *Issues in Accounting Education, 9*(2), 231–246.
Fairweather, J. S. (1996). *Faculty work and public trust: Restoring the value of teaching and public service in American academic life*. Needham Heights: Allyn & Bacon.
Fairweather, J. (2002). The mythologies of faculty productivity. *The Journal of Higher Education, 73*(1), 26–48.
Fairweather, J., & Beach, A. (2002). Variations in faculty work at research universities: Implications for state and institutional policy. *The Review of Higher Education, 26*(1), 97–115.
Feldman, K. A. (1987). Research productivity and scholarly accomplishment of college teachers as related to their instructional effectiveness: A review and exploration. *Research in Higher Education, 26*(3), 227–298.
Finkelstein, M. J. (1984). *The American academic profession: A synthesis of social scientific inquiry since world war II*. Columbus: Ohio State University.
Fox, M. (1985). Publication, performance, and reward in science and scholarship. *Higher Education: Handbook of Theory and Research, 1*, 255–282.
Fox, M. (1992). Research, teaching, and publication productivity: Mutuality versus competition in academia. *Sociology of Education, 65*, 293–305.
Garfield, E. (2006). The history and meaning of the journal impact factor. *JAMA, 295*, 90–93.
Golden, J., & Carstensen, F. V. (1992). Academic research productivity, department size and organization: Further results, comment. *Economics of Education Review, 11*(2), 153–160.
Hagerman, R., & Hagerman, C. (1989). Research promotion standards at select accounting programs. *Issues in Accounting Education, 4*(2), 265–279.
Harris, G. T. (1990). Research output in Australian university economics departments: An update for 1984–88. *Australian Economic Papers, 29*(55), 249–259.
Harry, J., & Goldner, N. (1972). The null relationship between teaching and research. *Sociology of Education, 45*, 47–60.
Hattie, J., & Marsh, H. (1996). The relationship between research and teaching: A meta-analysis. *Review of Educational Research, 66*(4), 507–542.
Hirsch, J. (2005). An index to quantify individual's scientific research output. In: *Proceedings of the National Academy of Science*. http://www.pnas.org/ content/102/46/16569.full. Accessed on 18 Dec 2009.
Hirsch, J. (2007). Does the h-index have predictive power? In: *Proceedings of the National Academy of Science.* http://www.pnas.org/content/104/49/19193.abstract Accessed on 18 Dec 2009.
Jordan, J., Meador, M., & Walters, S. (1989). Academic research productivity, Department size, and organization: Further results. *Economics of Education Review, 8*(4), 345–352.
Kyvik, S. (1995). Are big university departments better than small ones? *Higher Education, 30*(3), 295–304.

Levin, S., & Stephan, P. (1991). Research productivity over the life cycle: Evidence for academic scientists. *The American Economic Review, 81*(1), 114–132.

Linsky, A., & Straus, M. (1975). Student evaluations, research productivity, and eminence of college faculty. *Journal of Higher Education, 46*(1), 89–102.

Long, J.S. (1990). The origins of sex difference in science. *Social Forces, 64*, 1297–1316.

Marsh, H. W. (1987). Student evaluations of university teaching: Research findings, methodological issues, and directions for further research. *International Journal of Educational Research, 11*(3), 253–388.

Marsh, H. W., & Hattie, J. (2002). The relation between research productivity and teaching effectiveness. *Journal of Higher Education, 73*(5), 603–641.

Massy, W., & Wilger, A. (1995). Improving productivity: What faculty think about it and its effect on quality. *Change, 27*(4), 10–21.

McDaniel, E., & Feldhusen, J. (1970). Relationships between faculty ratings and indices of service and scholarship. In: *Proceedings of the 78th Annual Convention of the American Psychological Associations* (Vol. 5, pp. 619–620).

Menges, R. J. (1981). Instructional methods. In A. W. Chickering (Ed.), *The modern American college* (pp. 556–581). San Francisco: Jossey Bass.

Meyer, K. A. (1998). Faculty workload studies: Perspectives, needs, and future directions. *ASHE-ERIC Higher Education Reports, 26*(1), 1–129.

Middaugh, M. (2001). *Understanding faculty productivity: Standards and bench-marks for colleges and universities*. San Francisco: Jossey Bass.

Nettles, M. & Perna, L. (1995, November). Sex and race differences in faculty salaries, tenure, rank, and productivity: Why, on average, do women African Americans, and Hispanics have lower salaries, rank and tenure? Paper presented at the Association for the Study of Higher Education meeting, Orlando.

Porter, S., & Umbach, P. (2001). Analyzing faculty workload data using multilevel modeling. *Research in Higher Education, 42*(2), 171.

Quacquarelli Symonds. (2009). QS ranking of top universities. http://www.topuniversities.com/university-rankings/world-university-rankings/2009/regional-rankings/europe. Accessed on 15 Dec 2009.

Ramsden, P. (1994). Describing and explaining research productivity. *Higher Education, 28*, 207–226.

Ramsden, P., & Moses, I. (1992). Association between research and teaching in Australian higher education. *Higher Education, 23*, 273–295.

Rau, W., & Baker, P. J. (1989). The original contradictions of academe: Barriers facing the next academic revolution. *Teaching Sociology, 17*, 161–175.

Seglen, P. (1997). Why the impact factor of journals should not be used for evaluating research. *British Medical Journal, 314*(7079), 498–502.

Slaughter, S., & Rhoades, G. (2004). *Academic capitalism and the new economy: Markets, states and higher education*. Baltimore: Johns Hopkins.

Toutkoushian, R., Bellas, M., & Moore, J., III. (2007). The interaction effects of gender, race and marital status on faculty salaries. *Journal of Higher Education, 78*(5), 572–601.

Wardle, D. (1995). Journal citation impact factors and parochial citation practices. *Bulletin of the Ecological Society of America, 76*(2), 102–104.

Chapter 7
Measuring the Research Performance of Postsecondary Institutions

Robert K. Toutkoushian and Karen Webber

7.1 Introduction

Now more than ever, interest in measuring the research performance of colleges and universities is at an all-time high across the globe. There are several factors that have precipitated this growth of interest in assessing research productivity, particularly in the United States. First, colleges and universities are increasingly competing with each other for reputation and prestige, and enhancing research productivity is often viewed as a means to accomplish this goal. Based on expenditures for academic research, scientific production grew exponentially over the twentieth century (Geiger 2004) and current levels of research funding indicate the trend continues in the new century. As a result, the system of higher education in the United States has experienced considerable "mission drift" in recent years, as institutions that previously may have focused more on the teaching dimension of their mission have ratchetted up their research production and expectations of faculty. Some scholars including Geiger (2004) and Slaughter and Rhoades (2004) purport that universities today are highly reliant on federal and industry funding for research and development (R&D) funding, leading to "academic capitalism" and the possibility of research for financial gain more than for authentic discovery of knowledge. The benefits of funding and visibility that come from academic research are sought by many and factor into institution rankings as well. China's "2020 Plan," Korea's plan to create more world-class universities (Shin 2009), and the zealous grab for higher institutional rankings continues with annual releases of publications from such groups as *US News & World Report*, The Academic Ranking of World Universities (from Shanghi Jiao Tong University), and Thomson Reuters.

A second factor that has directed attention toward measuring research output of postsecondary institutions is what is known as the "accountability movement." Simply put, higher education stakeholders are increasingly asking for evidence that

R.K. Toutkoushian (✉) and K. Webber
Institute of Higher Education, University of Georgia, Athens, GA, USA
e-mail: rtoutkou@uga.edu

J.C. Shin et al. (eds.), *University Rankings*, The Changing Academy – The Changing Academic Profession in International Comparative Perspective 3,
DOI 10.1007/978-94-007-1116-7_7, © Springer Science+Business Media B.V. 2011

the resources being used for research are well-spent. In particular, the federal government has been a major benefactor in the production of research in the United States. For example, in 2005–2006, US postsecondary institutions received $30.3 billion in Federal grants and contracts alone (National Center for Education Statistics 2009). With this substantial level of investment, it is not surprising that the federal government is asking whether the investments in higher education research are paying off in terms of output (Shin 2009).

Especially in research universities, a substantial portion of federal support for postsecondary education is for research and development. Many institutions keep a close eye not only on their institution's level of research and development (R&D) funds, but also how their level of funding compares to peer institutions. Figure 7.1 shows an example of how university officials might examine the federal R&D funding for their institution relative to other schools. Institution officials identify a set of institutions with similar characteristics called *peers*. Officials will also identify a different set of institutions with characteristics (such as similarities in the number or type of students or faculty, facilities, and/or resource allocation) to which they strive in the future, called *aspirants*. Peer institutions enable officials to compare their institution with other schools who are similar at the time of comparison, whereas aspirant institutions provide benchmarks for policies and practices that will help an institution move to a higher level. As shown in the figure, the institution under study (called the *focus institution*) receives fewer federal funds for research and development than its peers throughout the 30-year time period. The gap in federal R&D expenditures between the focus institution and its aspirant institutions is even larger than the gap between the focus and peers. Not only does the focus institution receive less federal research money for R&D than its peers and aspirants, but also the gap in federal R&D expenditures has grown wider over time. This data would be important to institution officials who are examining fund allocation and the impact of allocation on the institution's research mission.

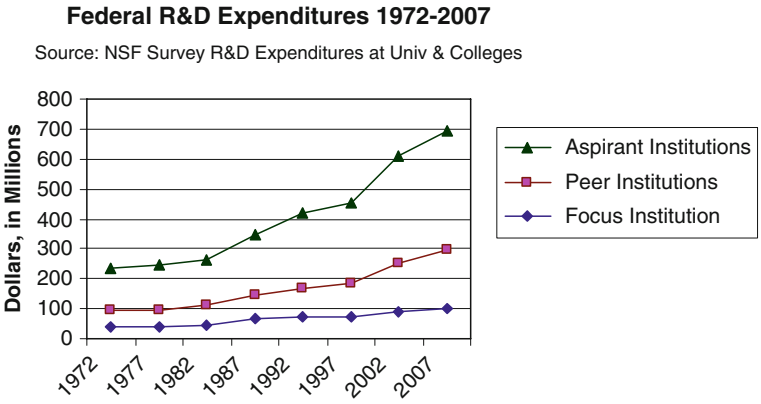

Fig. 7.1 Federal Research and Development expenditures for five institutions, 1972–2007 (Source: NSF Survey R&D Expenditures at University and Colleges)

Other higher education stakeholders are also contributing to the increased calls for accountability. For example, in 2005–2006, US postsecondary institutions received almost $66 billion from state governments (National Center for Education Statistics 2009) as well as substantial funding from private donors. Although the vast majority of financial support from state governments in the United States is not made for research support per se, state governments rely on research produced by their institutions to help improve the quality of life for their citizens. For example, in FY 2007, the state of Georgia apportioned $1.9 billion to state-related postsecondary institutions (USG Annual Financial Report FY 2007). Private individuals also help support the research infrastructure of postsecondary institutions through their donations. For example, private gifts totaled $5 billion in 2005–2006 (National Center for Education Statistics 2009).

Finally, another reason for the importance of examining an institution's research productivity relates to the fact that authentic knowledge production is important to the economic and social development of states and nations. Research has the potential to lead to developments that will improve the standard of living and quality of life for citizens. A number of scholars have examined the connection between research and economic growth (Etzkowitz 2008; Geiger 2004; Geiger and Sa 2008). Etzkowitz and Dzisah (2008) believe that knowledge-based development is rooted in the university as the institution that generates human and intellectual capital. Etzkowitz's triple helix model asserts the university as a third critical element in economic organization and development, and as we move from an industrial to a knowledge-based society, universities will play an even greater role in innovation and economic development. Research is also thought to contribute to other goals of countries such as enhancing their national security (Libaers 2009; James 2009).

Nonetheless, the increased calls for measures of research productivity have been hampered by the fact that it is difficult to accurately measure the research production of institutions of higher education. Despite technological advances that enable policy makers to tabulate information on research in more efficient ways than were possible a decade earlier, the state of the art in measuring an institution's research productivity remains fairly primitive. The challenges encountered when attempting to do this fall into one of the following three categories: (a) difficulties in obtaining data, (b) distinguishing between the quality and quantity of research produced, and (c) difficulties in aggregating research productivity across fields/disciplines.

In this chapter, we focus on the many aspects related to the measurement of institutional research productivity in academe. We begin by reviewing the context in which the measurement of institutional research productivity occurs. We then describe the ways in which research productivity is commonly measured around the world, and discuss the advantages and disadvantages of each approach. Finally, we offer some observations on future directions for measuring institutional research productivity. Although much of the discussion in this chapter is focused on colleges and universities in the United States, the issues that we describe apply equally to institutions around the globe.

7.2 The Context for Measuring Institutional Research Performance

Institutions of higher education are incredibly complex organizations. The activities of colleges and universities are typically classified according to whether they are related to the research, teaching, or service aspects of their mission. Some colleges may specialize in only one of these areas, while others seek to make contributions in all three areas to varying degrees. It is not clear, the extent to which the research, teaching, and service missions of colleges and universities compete with each other for resources or assist each other in production (Dundar and Lewis 1995). It is perhaps best to think that both are true: Time spent in any one activity takes away time from the other two activities, and yet there can be complementarities between them (Becker 1975, 1979, 1982). For example, a faculty member may use her research to enhance her teaching by bringing more current examples into the classroom, and it can also improve her service to local, state, and national organizations. The tripart mission of institutions of higher education is important to understand when looking at research production because variations in research productivity could be due to differences in the emphases given to teaching and service (Dolan and Schmidt 1994; Porter and Toutkoushian 2006). This difference in emphasis can subsequently affect the aggregated amount of research produced by the institution.

The way in which activities are produced within a college or university is also very unique. Research, teaching, and service take place in separate academic departments or colleges, which largely operate independently of each other. Furthermore, most of the production of these outputs is accomplished by individual workers (e.g., faculty members) who often function as independent contractors with little direction and coordination by the employer. Even though most faculty and departments are involved in producing research, teaching, and service, the specifics of their work varies greatly across departments and even faculty within a department. This has been discussed at length by many including Baird (1986), Becher (1994), Biglan (1973), and Marsh and Hattie (2002). Accordingly, the type of research typically generated by a faculty member can look quite different depending on the department in which the individual is employed. This is important for the measurement of institutional research productivity because this quantity requires the aggregation of research across individuals and departments within the institution. Variations in the emphasis of different academic fields can, therefore, contribute to differences in research output at the institutional level.

Another complicating factor is the difficulty in determining what is meant by research productivity. The term "productivity," as opposed to output, implies a measurement of output per unit of input. This follows from the production function model of economics, where inputs are converted into outputs through the production process used by an organization. A similar description can also be found in the I-P-O model in education (Astin 1970; 1991). This is consistent with the notion of efficiency, where resources are being used optimally to produce research output.

This means that if one is going to try to compare organizations on the basis of their productivity, it is essential to adjust the total output produced by the quantity of inputs and production processes used. In our example, it would not be correct to conclude that the University of Minnesota is more productive in research than Carleton College solely because the former produces more publications in a given year, if the difference in output is due to the University of Minnesota having more faculty members and financial resources to produce research.

7.3 Challenges in Measuring Institutional Research Performance

There are a number of empirical challenges that must be addressed when attempting to measure the research performance of an institution. These can be grouped into three categories: (a) lack of available data, (b) difficulty in measuring the quality versus quantity of research, and (c) difficulties in aggregation. We will discuss each in turn.

7.3.1 Lack of Available Data

As is true with most empirical study, analysts are restricted by the data at their disposal. This certainly holds when attempting to measure an institution's research productivity. Ideally, we should be able to report information on the benefits to students and society through a college's or university's research activities and use this to assess their performance. Unfortunately, we rarely have a subset of data that clearly relates research activities to outcomes or benefits. What we can observe, however, are metrics such as dollars expended for research and dollars received through research grants. For this reason, the old Carnegie classifications of institutions by research intensity relied almost exclusively on research dollars received through grants and the number of graduate degrees awarded, on the premise that research-intensive institutions are more heavily involved in graduate education. Other data that are more difficult to obtain than research dollars, but can be collected, include counts of publications (Toutkoushian et al. 2003), patents, and citations. Although this is a positive step away from resources used to produce research and the output generated by research, the quantities still may not fully capture the quality of research produced by members of the institution.

In an attempt to shift our thinking on academic scholarship, Boyer (1990) argued that colleges should restructure activities (and therefore outcomes that result) around four dimensions of scholarship. Revising current activities within these four dimensions offers a different way to think about and quantify measure of productivity, and would include more elements related to teaching and service.

While Boyer's ideas have generated interest in academic circles, we see little evidence that this has had an effect on how institutional research productivity is measured, due, largely, to data limitations. It is easy for analysts to envision current measures of production such as publications, patents and citations, and more difficult to determine how to accurately measure certain aspects of teaching and efforts devoted to institution and community service.

7.3.2 Quality Versus Quantity

An institution's research production can be described in terms of the *quantity* of research produced or the *quality* of research produced. In some arenas, the total level of extramural funding a faculty member receives in a period of time is used to gauge quality (more funding signals higher quality), but in many cases the total amount of funding does not necessarily indicate quality. There are some attempts to differentiate research output by quality, for example, counts of peer-reviewed separate from non-peer-reviewed as well as different levels for journals by a set of criteria to designate rigor. These efforts are not used widely, although, are more frequent in some disciplines such as business and law. Although one would think that the optimal construct of institutional research productivity should reflect the quality of research produced, most currently used metrics are based on the quantity of research produced. As with Boyer's (1990) attempts for the four dimensions of scholarship, analysts and other officials within the academy struggle to find accurate metrics that truly capture quality, and thus often return to the quantitative measures that are more familiar to researchers and policymakers.

7.3.3 Aggregation

As noted earlier, institutions of higher education are organized around individual academic departments that can vary greatly in terms of the type of research that they conduct, and the way in which their research is translated into outputs. Although in many fields research is evaluated based on publications in peer-reviewed academic journals, this does not apply across-the-board. In some disciplines such as history, research is more frequently communicated via books and not journal articles (Huang and Chang 2008; Moed 2005, 2006; Nederhof 2006; Tang 2008; Zainab and Goi 1997), and neither books nor journal articles may be appropriate for measuring research productivity in other fields such as the performing arts. This raises the question of what should be aggregated across departments in order to represent the research output of an institution: Is it journal articles, books, patents, citations, or something else? If multiple items are to be aggregated, how should they be weighted? The aggregation problem also raises the issue that any output measure that is aggregated across departments will be affected by the disciplinary

mix at an institution. For example, if hard science departments are more likely than departments in the humanities to produce research in the form of articles in peer-reviewed journals, then an institution with a greater emphasis on the humanities would appear to be less productive than other institutions using this metric.

7.4 Current Approaches to Measuring Institutional Research Productivity

There are a number of approaches that have been used to measure the research productivity of postsecondary institutions. For more discussion on the issues relating to different approaches for measuring institutional research activities, see Diamond and Graham (2000), Johnes et al. (1993), and Tan (1986). Due to the availability of data, the most common approaches currently in use rely on either input or output measures of research and emphasize quantity over quality.

7.4.1 Research Funding

Perhaps the most frequently used barometer of research productivity for colleges and universities is the amount of funding received or expended for research. This can be thought of as an input into the production of research because in many fields, funding is needed to purchase equipment, supplies, staff, and faculty time to conduct research, and it is reasonable to assume that there would be a causal relationship between research funding and the production of research. Even though the amount of research funding is an input into production, it could be argued that research funding has a quality aspect to it as well because research dollars are usually awarded on a competitive basis. Accordingly, receipt of a research grant is an indicator of the quality of the research proposal developed by the principal investigators. Part of the popularity of grant funding as a measure of research productivity is that it is readily available for most colleges and universities through their standard financial reports.

Even though many entities use funding expenditures as a measure of institutional research productivity, there are reasons why this may not be the ideal measure. First, the reliance on grant funding can vary substantially across fields within an institution. Although laboratory science departments such as chemistry and engineering may require grant funding to purchase expensive equipment needed to conduct research, the same is generally not true in the humanities, arts, and some social sciences. This implies that the disciplinary mix of an institution can greatly affect the amount of research funding it can generate, and that there is not a one-to-one connection between the amount of grant funding received and the quality/quantity of research produced. It should also be noted that not all grant

funding is awarded on a competitive basis, which adds further variation into the relationship between grant funding and research productivity.

7.4.2 Publication Counts

A more recent trend in academia – especially internationally – has been to use bibliometric methods (publication or citation counts) to measure an institution's research productivity. The publication count for an institution would represent the number of publications in specific categories produced by individuals who are affiliated with the institution. Such publication counts are appealing to researchers because they are an output-based measure as opposed to an input-based measure as is true for grant funding. Publication counts also have a quality dimension imbedded in them that most articles and books must pass through some form of peer review prior to being published. Therefore, the act of being published reflects the assessment of some experts in the field that the research will make a contribution to the literature. This quality dimension of publication counts is magnified by the fact that publishing is a competitive process due to space limitations in academic journals. Therefore, some articles that might be deemed worthy of publication are turned down in favor of those of even higher perceived quality.

Despite their appeal, publication counts have not been used extensively in the past to measure research productivity because information on publication counts was difficult to find. There was no central source where one could go to count the number of publications produced by individuals at an institution. However, with the advent of databases and electronic access to these databases, it is much more feasible today than it was 10 years ago to obtain publication counts for a range of institutions (Toutkoushian et al. 2003). Newer electronic databases such as the Web of Knowledge produced by the Institute of Scientific Information (ISI, now Thomson Reuters) make it even easier for researchers to search on publications produced by individuals based on their institutional affiliation.

There are, however, a number of important limitations with publication counts as a measure of institutional research output. As described earlier, fields vary greatly in terms of their reliance on publications in academic journals as a means for disseminating knowledge from research. Accordingly, variations in publication counts across disciplines could reflect the norms for the field and not the quality of the research per se. Some fields value books and technical reports, but many citation databases (such as ISI Thomson Reuters) do not include books and most conference proceedings. This aggregation problem reduces the validity of these counts as measures of total research productivity. Second, many published works are never cited and have little impact on the field or society at large. Third, when conducting international comparisons of institutions on publication counts, it is likely that such counts would favor authors affiliated with institutions in the same geographic region as the journal published (Wardle 1995). This bias could be due to the proliferation of English-language journals in academe.

Finally, there are a number of data issues that are important when examining publication count data as a measure of research productivity. The ISI databases are limited in the scope of journals that they monitor, and thus a number of publications will be overlooked if they did not appear in the journals being tracked. On the other hand, it could be argued that the publication counts from journals tracked by Thomson Reuters are a better measure of quality because these journals tend to be the more well-known journals in their respective fields. A second issue is that there can be inaccuracies in the database due to misspellings of institutional name, incorrect author affiliations, and incorrect citations in the original papers. This problem becomes magnified when taking the analysis down to the level of an individual faculty member, where variations on the spellings of names can greatly impact an individual's publication counts. Care has to be taken when aggregating articles with multiple authors to ensure that an article with three co-authors from a single institution is only counted once. Finally, one has to decide whether non-journal articles should also be counted, and what weights should be attached to them. Some work suggests that a substantial amount of cited research is in the form of monographs and/or books (Al et al. 2006; Krampen et al. 2007; Porta et al. 2006; Stern 1983; Yates and Chapman 2005; Zainab and Goi 1997). Nonetheless, it can be argued that as a rough measure of institutional research productivity, the publication counts obtained from these databases are a fairly good approximation of total productivity and are relatively easy to obtain.

7.4.3 Citation Counts

Another possibility for measuring institutional research productivity is to count the number of citations attributed to individuals at an institution. Because the Thomson Reuters database tracks the citations to the articles that they cover, it is more feasible than ever to sum the citations for articles written by individuals at an institution. The primary advantage of citation counts over publication counts and research dollars is that the counts arguably capture both the quantity and quality of research produced. Total citations to an institution should be positively related to the quantity of publications because there would be more publications to potentially cite. In addition, the act of being cited is an indicator that the research produced has been used by someone in their work, and thus could reflect the quality of the work produced. There is a sizable literature dating back more than 40 years, including, Cronin (1984), DeBellis (2009), Diamond (1986), Garfield (1979), Gerrity and McKenzie (1978), Laband (1985), Lindsey (1989), Moed et al. (1985), and Seol and Park (2008) that focus on the reasons why research is cited and how citation patterns emerge.

Researchers must decide on the best way to combine and report citations for use as a measure of institutional research productivity. Perhaps the most straightforward way of doing this is to count the number of citations that were received in a given year for all publications where one or more author's affiliation was the

institution in question. For example, one could total the citations received in 2009 to all articles published by individuals affiliated with one institution or a set of institutions within a state system. The citation counts would therefore apply to all articles published prior to 2009, and thus represent the "current value" of research published to date (Diamond 1986; Toutkoushian 1994). Such data could be used within one institution or to quasi-compare across institutions if so desired. Alternatively, one could track the citations received by publications in a given year. This would be more time-consuming and because of the time lag involved in citations, one would have to wait several years before being able to accurately measure the citations received by the publications. As a practical matter, it is usually best to use current citations to all previous work as a measure of an institution's research productivity.

Despite their potential value as a measure of an institution's research productivity, citations have several important limitations. First, as is true with grants and publication counts, citation frequencies can vary dramatically by field. This will create the same kind of aggregation problem as described earlier when attempting to use citation counts to compare institutions. Second, not all citations are a positive indicator of quality. In some instances, a researcher will cite another researcher's work to point out a flaw or error in their work. A counterargument, however, is that even a "bad citation" such as this has made a contribution to knowledge production by leading to another development. Third, the citations received by a publication will vary with the age of a paper (Cole 1979; Diamond 1984; Shin and Putnam 1982); so, the timing at which citations are counted can be important. Typically there is a lag between the publication of a paper and the time at which the paper can be cited. The lag occurs because of the time needed for researchers to read a study and incorporate it into their work, and then the time needed for the new work to be reviewed, accepted for publication, and appear in print. Accordingly, it would not be unusual for a high-quality research paper to receive no citations for the first few years after publication. Likewise, there will probably be a point at which the number of citations received by a paper begins to decrease as newer research takes precedence in the field. Fourth, the citation counts received in a given year to all prior publications is backward-looking in that, the productivity measure is related to research that was completed in the past. Due to the lag in citations described here, however, it is very difficult to devise an alternative that would capture the value of more current research unless the counts were restricted to articles published within a range of years. Finally, as mentioned above, it is possible that articles published in one region (say North America versus Europe) would include more citations from those in the same region (North American authors) and/or English-language journals would receive more citations than comparable articles in non-English language journals, thus biasing the citation counts against institutions that are not in the same proximity or where the predominant language of researchers is not English (Seglen 1997).

There are also a number of potential measurement issues that can be important when looking at citations as a measure of institutional research productivity. As is true with publications, there can be errors in the citations made by authors that would have an impact on citation counts. For example, if an article is not properly

cited by an author, it is possible that the citation will not be attributed to the correct institution and this could be missed when counting citations. Researchers also need to ensure that an individual's citations to co-authored papers are not counted multiple times (Long and McGinnis 1982; Roy et al. 1983). Perhaps, the biggest technical problem that must be addressed with citation counts is how to handle self-citations. The citations received by a paper may include some citations made by the authors of the original paper. This practice is not unusual because researchers typically build upon their own work and may rightfully want to call this to the attention of readers. However, self-citations introduce the possibility that the counts could be inflated and overrepresent the extent to which others in academe have found the research to be useful and valuable.

One additional issue related to citation counts includes the recent inclusion of new citation databases. Google Scholar and Scopus are recent newcomers, and a number of scholars have compared results across ISI, Google Scholar and Scopus. For example, Bauer and Bakkalbasi (2005) compared citation counts from World of Science, Scopus, and Google Scholar in 1985 and 2000 for the field of information sciences. They found that World of Science provided the highest citation counts for 1985 articles and Google Scholar had a much higher count for 2000. In another analysis for the two disciplines, Bakkalbasi (2006) found that each of the three databases yielded some unique citations, but also that the differences in citation counts varied between the two fields. These authors conclude that the most complete set of citations may depend on the subject and publication year of a particular article.

7.4.4 Professional Award Counts

Another potential measure of an institution's research productivity can be derived by counting the number of faculty who have received professional awards for their scholarship. There are entities that bestow recognition across a range of disciplines, the most famous of which is the Nobel Foundation in Sweden that annually awards the Nobel Prize to faculty in chemistry, physiology/medicine, physics, and (through a related entity named Sveriges Riksbank) economic sciences, and also awards the Nobel Peace Prize and the Nobel Prize for literature. Table 7.1 shows the 11 universities through 2009 with 10 or more faculty members who have been awarded a Nobel Prize in chemistry, physics, physiology/medicine, or economic sciences. Of these 11 institutions, Harvard University is by far the world leader with 31 Nobel Laureates, and 9 of the 11 institutions are in the United States.

The National Academies of Sciences (http://www.nasonline.org) recognizes faculty for their scholarly achievement in a number of science-related fields including physics, biology, medicine, and social sciences. Election to the National Academy of Sciences is considered to be one of the highest honors that can be bestowed on a scientist or engineer. Begun in 1863, The National Academy of Sciences, a society of distinguished scholars, engaged in scientific and engineering research, dedicated to the furtherance of science and technology and to their use for

Table 7.1 Institutions with ten or more Nobel Laureates through 2009

Institution	Number of Nobel Laureates
Harvard University	31
Stanford University	18
Massachusetts Institute of Technology	17
California Institute of Technology	16
Columbia University	16
Max Planck Institute (Germany)	16
Rockefeller University	16
University of California at Berkeley	16
University of Cambridge (United Kingdom)	16
University of Chicago	16
Princeton University	11

Source: Nobel Foundation

the general welfare. There are approximately 2,100 members and 380 foreign associates in the NAS, many of whom have been awarded Nobel Prizes. In addition, most individual fields have their own awards that they grant to the distinguished faculty. These would include the Wolf Prize (chemistry), the Pulitzer Prize (journalism), the Leroy P. Steele Prize (mathematics), the Enrico Fermi Award (energy science), the National Academy of Engineering, and the National Medal of Science, to name but a few.

Although appealing, counts of faculty award recipients can be very difficult to use as a reliable measure of an institution's research productivity. Decisions have to be made about which awards should be counted, and whether all awards should all be given the same weight (e.g., does having a Nobel Prize winner in economic sciences count more than having a faculty member elected to the National Academy of Sciences?). The criteria for receipt of an award can be more stringent for some awards than for others. Data on the institutional affiliations of award recipients, at the time of their award, is likely to vary greatly. The mix of disciplines at an institution can obviously affect the resulting counts of award winners depending on which awards are included in the totals. For example, the counts of Nobel Prize winners will obviously favor institutions with a strong emphasis on hard sciences. Counts of award winners would also be biased towards larger institutions with more faculty members, unless the counts could be expressed on a per-capita basis. An even more fundamental question is whether the receipt of an award is a valid measure of current research productivity, or reflects past productivity. For this reason, counts of award recipients would most likely favor institutions with a more senior faculty.

7.4.5 Patent Counts

Finally, counts of the patents produced by an institution can be used as a measure of research productivity. Patent counts can be appealing because, in theory, they

reflect the practical application or use of research. The counts also in part reflect the quality of research because of the review that is part of the patenting process.

However, patent counts have several limitations as a measure of research productivity. First, not all research is appropriate for patents. Second, patent counts would tend to advantage institutions that emphasize fields where research can more easily lead to patents, such as the hard sciences and engineering. Third, not all patents lead to improvements in society; researchers can obtain patents for their work for any of a number of reasons. Finally, data on patents is often more difficult to obtain than is true for other measures of research.

7.5 Entities Involved in Measuring Institutional Research Productivity

There are a couple of main groups that have been involved in measuring the research performance of institutions of higher education. These groups differ in terms of the information that they collect and the reasons for the approaches that they use. Due to space limitations, we only review several of the more prominent entities in this chapter. The Education and Social Science Library at the University of Illinois at Urbana-Champaign (http://www.library.illinois.edu/edx/rankint.htm) and the Institute for Higher Education Policy (http://www.ihep.org/Research/rankingsystemsclearinghouse.cfm) provide more information on many of the ranking systems that currently exist both in the United States and internationally.

7.5.1 Carnegie Commission on Higher Education

Perhaps, the foremost group in the United States that has been involved in measuring the research produced by institutions of higher education is the Carnegie Commission on Higher Education. The Carnegie Commission is best known for their classification scheme of colleges and universities, where institutions were placed into one of several categories depending on the amount of research funding that they receive and the number of graduate degrees awarded. The first Carnegie classification system was developed in 1970 (Carnegie Commission on Higher Education 1971), and the system has been revised in 1976, 1987, 1994, 2000, and 2005. Table 7.2 shows the categories that were put in place in 1994 and the criteria used to assign institutions to each category.

The Commission's stated purpose behind the categorization scheme was not to rank institutions, but rather to help inform research on higher education by enabling researchers to identify roughly comparable groups of institutions (McCormick and Zhao 2005). During the 1990s, however, the Carnegie Commission on Higher Education became increasingly concerned that their categorization scheme was

Table 7.2 Carnegie classification scheme for 4-year institutions, 1994

Category	Minimum annual federal grant support	Minimum annual degree production
Research Universities I	$40 million	50 doctorate degrees
Research Universities II	$15.5 million	50 doctorate degrees
Doctoral Universities I	No minimum	40 doctorate degrees across at least five disciplines
Doctoral Universities II	No minimum	10 doctorate degrees across at least three disciplines, or 20 doctorate degrees total
Masters Colleges and Universities I	No minimum	40 masters degrees across at least three disciplines
Masters Colleges and Universities II	No minimum	20 masters degrees overall
Baccalaureate Colleges I	No minimum	40% of bachelors degrees in liberal arts fields, and restrictive admissions
Baccalaureate Colleges II	No minimum	Less than 40% of bachelors degrees in liberal arts, or less restrictive in admissions

being used to rank institutions, and that some institutions were attempting to make changes in their graduate degrees awarded and/or research expenditures in order to move up to the next category. The Commission has been very explicit about its intended purpose for categorizing institutions, stating on their website:

> The Carnegie Foundation does not rank colleges and universities. Our classifications identify meaningful similarities and differences among institutions, but they do not imply quality differences. (Carnegie Commission on Higher Education 2009).

To address their concerns, in 2000, the Carnegie Commission on Higher Education collapsed the four research categories (Research I, II and Doctoral I, II) into two categories: Doctoral/Research Universities – Extensive, and Doctoral/Research Universities – Intensive. The stated reason behind the reorganization of the categories was described as follows by the Commission:

> The Research I & II and Doctoral I & II categories of doctorate-granting institutions last appeared in the 1994 edition. The use of Roman numerals was discontinued to avoid the inference that the categories signify quality differences. The traditional classification framework has been updated and is now identified as the Basic Classification. Many of the category definitions and labels have changed with this revision. (Carnegie Commission on Higher Education 2009)

Finally, amidst continuing concern over what was perceived to be the misuse of their classification system, the Carnegie Commission in 2005 again changed their system to one where multiple classifications are used for institutions according to their instructional programs, profile of enrolled students, and size and residential status. The new classifications provide users a different way to view US colleges and universities, offering researchers greater flexibility in meeting their analytic needs, and do not place as much emphasis as before on the total levels of research funding.

7.5.2 The Center for Measuring University Performance

Perhaps, the most explicit effort to quantify the research productivity of institutions in the United States and rank them on this basis is produced by The Center for Measuring University Performance (http://mup.asu.edu/). The Center, as it is more commonly known, originated at the University of Florida and is now located at Arizona State University. The first set of institutional rankings by The Center was compiled in 2000, and the rankings have since been updated annually. In contrast to the Carnegie Commission on Higher Education's desire to avoid having their classifications used to rank institutions, The Center's goal is clearly to rank institutions for policy makers.

The rankings produced by The Center are based on nine categories:

- Total research expenditures
- Federal research expenditures
- Endowment assets
- Annual giving
- National Academy of Sciences members
- Number of faculty awards
- Number of doctorate degrees granted
- Number of postdoctorates
- Range of SAT scores for undergraduates

The Center compiles data on these nine elements and then ranks institutions on the basis of each category. The final ranking for each institution is determined by the number of these nine categories in which an institution is ranked in the top 25. For example, in the 2008 rankings, three institutions – Columbia University, Massachusetts Institute of Technology, and Stanford University – were tied for the top ranking because they were ranked in the top 25 on all nine criteria. Another five institutions (Harvard University, University of Pennsylvania, Yale University, Duke University, University of Michigan) were ranked in the top 25 in eight of the nine categories and thus were tied for fourth overall in The Center's rankings.

The Center's philosophical approach to ranking institutions is also deliberately void of statistical analysis. Rather, The Center believes that there is value in presenting the raw data on various components to readers and using a very simple algorithm – the sum of top 25 rankings – as the means for ranking institutions. As noted by The Center (Capaldi et al. 2008: 2):

> A primary principle of this enterprise has always been and remains the production of nationally available data, compiled in a standardized format, made available free online for further analysis by colleagues, and presented without elaborate statistical manipulations.

A review of the methodology used by The Center reveals that only two of the nine categories – total and federal research dollars – are directly related to current research activities, and two others (NAS members and number of faculty awards) are related to academic reputation from the past and current research. At the same time, metrics such as the range of SAT scores and the annual giving to the institution

are only tangentially related to research productivity. Accordingly, despite its label, the rankings produced by The Center reflect a wider range of activities than simply research. Because the measures are not standardized by the number of faculty or size of the institution, the resulting rankings will naturally favor larger institutions. The Center is critical of international rankings of institutions that are based on bibliometric data such as publication or citation counts because "…the validity of linking publication and citation counts by author to the distinction of universities is questionable." (Capaldi et al. 2008: 3). Interestingly, The Center criticizes bibliometric counts on the grounds that they "…often misrepresent quantity for quality" even though this criticism applies equally to two of the metrics used by The Center (total and federal research expenditures).

7.5.3 The Centre for Science and Technology Studies (CWTS)

The Centre for Science and Technology Studies at Leiden University (http://www.socialsciences.leiden.edu/cwts/) is one of the leading entities in compiling international rankings of universities. The "Leiden Rankings," as they are more commonly known, date back to 2003 and consist of the 100 and 250 largest universities in the world (http://www.cwts.nl/ranking/LeidenRankingWebSite. html). The rankings produced by CWTS rely on the ratio of citations to publications over a specified number of years. The Leiden Rankings are made based on an institution's (a) total publications, (b) ratio of citations to publications, (c) ratio of citations to publications after adjusting for the size of the institution, and (d) total publications multiplied by the "university's field-normalized average impact." The Leiden Rankings are notable for their broad scope, encompassing a large number of institutions across the globe. Although the Leiden Rankings recognize the importance of controlling for scale when comparing institutions, CWTS provides little information about the methodology used to compute the field-normalized average impact, nor the rationale behind the specific adjustment as a better way to compare institutions. Also, as noted earlier, the reliance on bibliometric research measures may lead the rankings to favor institutions where faculty members are more fluent in English.

7.5.4 Other Bibliometric-Based Rankings of Institutions

In addition to CWTS, there are other organizations that compile international rankings of universities based on publication and/or citation data. The Higher Education Evaluation and Accreditation Council of Taiwan, for example, ranks 500 universities across the globe on the basis of publications and citations (current and total, for the past 11 years) as well as other criteria. The Graduate School of Education at Shanghai Jiao Tong University also produces rankings for the top 500 worldwide institutions, known as the "Academic Ranking of Worldwide Universities"

Table 7.3 ARWU criteria and weight used in the world universities rankings

Criteria	Indicator	Code	Weight (%)
Quality of education	Alumni of an institution winning Nobel Prizes and Fields Medals	Alumni	10
Quality of faculty	Staff of an institution winning Nobel Prizes and Fields Medals	Award	20
	Highly cited researchers in 21 broad subject categories	HiCi	20
Research output	Articles published in Nature and Science	N&S	20
	Articles indexed in Science Citation Index-expanded, and Social Science Citation Index	PUB	20
Per capita performance	Per capita academic performance of an institution	PCP	10
Total	0	0	100

(http://www.arwu.org/). Of the six components that they use to derive the rankings (see Table 7.3 for the components and weighting of each component), however, only two of them – total publication counts and counts of articles in the journals *Nature* and *Science* – are directly related to research production.

7.5.5 National Research Council Survey of Doctoral Programs

Although not an entity that focus on institutional-level research productivity, the National Research Council's (NRC) Survey on Assessment of Doctorate Programs (http://sites.nationalacademies.org/PGA/Resdoc/PGA_044475) is a useful source of information on the research produced by institutions in the United States. The National Research Council, which has previously conducted surveys of doctoral programs in the United States in 1982 and 1993 (with new results due to be released in 2010), collects information at the program level on faculty productivity (research publications), general institutional data, and perceptions about the scholarly quality of academic programs. Previous survey results have enabled NRC officials to create rankings that provide comparisons of doctoral programs in a field of study. The next iteration of survey results will enable analysts to perform additional analyses to more deeply compare their doctoral program with others.

As shown by Webster and Skinner (1996), it is possible to aggregate the program-level information collected in the NRC surveys to evaluate and rank institutions. For example, Webster and Skinner (1996) ranked institutions in the United States on the basis of the mean score of the scholarly quality of faculty for graduate programs covered in the NRC survey. There are, of course, limitations to the use of NRC data for assessing institutional productivity and quality (Webster and Skinner 1996; Maher 1996). Such rankings would tend to be biased towards larger institutions and those with a wider range of academic programs. The subjective nature of

the peer assessments calls into question whether what is being measured is actually productivity or reputation. Likewise, not all institutions and doctoral programs are included in the NRC survey. Finally, as discussed by Maher (1996), care must be taken when aggregating program-level ratings to the institutional level.

7.6 Summary and Discussion

The measurement of institutional research productivity is an enduring and growing topic in higher education. The growth can be attributed to the convergence of several factors: the rising emphasis on rankings and prestige in higher education, the accountability movement in all levels of education, and technological advances that have made bibliometric data on institutions easier to obtain and analyze.

Nonetheless, our review of the state of the art in measuring institutional research productivity reveals that the field is still fairly primitive. The first limitation in the field is that the various entities that use information on institutional research rely on a wide range of ad hoc approaches to combining data, and the approaches often have little stated justification. Why, for example, would the method used by The Center of basing rankings on the counts of individual rankings in the Top 25 be a better approach than a more statistical approach to examining research productivity? A second limitation is that most of the approaches do not control for the scale of operations. This is crucial if the rankings and evaluations are to be used for accountability purposes to examine research productivity as opposed to research output. As noted earlier, productivity implies a measure that shows output per unit of input. In their current state, however, most of the research metrics in use are greatly affected by the size of the institution and thus are biased towards larger institutions which have more faculty and staff to produce research. Finally, some of the rankings in use rely on a mix of metrics, only some of which are directly related to research production. Accordingly, the rankings that they obtain may not reflect research productivity per se. Policy makers need to be aware of these limitations of the approaches that are currently used to assess institutional research productivity.

As for future directions, we have several recommendations:

- *Standardize measures used in rankings*. It is crucial that whatever metrics are being used, they should be expressed per unit of input if the goal is to derive measures of research productivity. One way is to simply divide specific metrics by the number of students or faculty at an institution, or total revenue. Regardless, a rationale needs to be given for the standardization used, and perhaps tests of robustness conducted to see how sensitive rankings are to different approaches of standardization.
- *Focus on research components*. Given that there are several readily-accessible metrics related to research available to analysts, it would be helpful to see an initiative that would attempt to pull all of these together into a single analysis.

- *Provide rationale for ranking method used*. It is not sufficient for entities to use an algorithm for ranking institutions without providing analysis and justification for their approach. This is particularly important because the resulting rankings and evaluations may depend upon the algorithm used.
- *Consider joint production of research, teaching, and service*. Finally, it would be informative to see how the other two main aspects of institutional activities – teaching and service – influence research production. This may require an analysis that examines the interrelationship among outputs (Porter and Toutkoushian 2006).
- *Study connection of research production to long-term outcomes.* In order to move beyond simply counting what we can count, studies are needed that will examine how research production is related to desired outcomes from higher education. Is there evidence that research leads to economic gains for states and nations? To what extent is research used by the industry to develop new products and services that can benefit society? Are students better off in terms of employment as a result of university research? All of these questions are very important, and would require special research designs that could connect current research activities to future benefits.

Consider use of new information sources to measure research productivity. Finally, there are new technologies available with information that could be connected to research output and perhaps productivity. It is now possible to conduct a search on individuals via the Internet and possibly find new measures of how their work is being used.

References

Al, U., Sahiner, M., & Tonta, Y. (2006). Arts and humanities literature: Bibliometric characteristics of contributions by Turkish authors. *Journal of the American Society for Information Science and Technology, 57*(8), 1011–1022.

Astin, A. (1970). Methodology of research on college impact (I). *Sociology of Education, 43*(3), 223–254.

Astin, A. W. (1991). *Assessment for excellence*. San Francisco: Jossey Bass.

Baird, L. L. (1986). What characterizes a productive research department? *Research in Higher Education, 25*(3), 211–225.

Bakkalbasi, N., Bauer, K. Glover, J., & Wang, L. (2006). Three options for citation tracking: Google Scholar, Scopus, and Web of Science. *Biomedical Digital Libraries*. Retrieved January 27, 2010 at, http://www.ncbi.nlm.nih.gov/pmc/articles/PMC1533854/pdf/1742-5581-3-7.pdf

Bauer, K., & Bakkalbasi, N. (2005). An examination of citation counts in a new scholarly communication environment. *D-Lib Magazine*. Retrieved January 27, 2010 at http://www.dlib.org/dlib/september05/bauer/09bauer.html

Becher, T. (1994). The significance of disciplinary differences. *Studies in Higher Education, 19*(2), 151.

Becker, W. (1975). The university professor as a utility maximizer and producer of learning, research, and income. *The Journal of Human Resources, 14*(1), 109–115.

Becker, W. (1979). Professorial behavior given a stochastic reward structure. *The American Economic Review, 69*(5), 1010–1017.
Becker, W. (1982). Behavior and productivity implications of institutional and project funding of research: Comment. *American Journal of Agricultural Economics, 63*(3), 595–598.
Biglan, A. (1973). Relationships between subject matter characteristics and the structure and output of university departments. *The Journal of Applied Psychology, 57*(3), 204–213.
Boyer, E. (1990). *Scholarship reconsidered: Priorities of the professoriate*. Princeton: Carnegie Foundation.
Capaldi, E., Lombardi, J., Abbey, C., & Craig, D. (2008). *The top American research universities: 2008 annual report*. Tempe, AZ: The Center for Measuring University Performance. Retrieved October 15, 2009, from http://mup.asu.edu/research2008.pdf
Carnegie Commission on Higher Education. (1971). *New students and new places*. New York: McGraw-Hill.
Carnegie Commission on Higher Education. (2009). Carnegie classifications FAQs. Retrieved October 15, 2009, from http://www.carnegiefoundation.org/about/sub.asp?key=18&subkey=405#1.0.1
Cole, S. (1979). Age and scientific performance. *The American Journal of Sociology, 84*(4), 958–977.
Cronin, B. (1984). *The citation process: The role and significance of citations in scientific communication*. London: Taylor Graham.
DeBellis, N. (2009). *Bibliometrics and citation analysis: From the science citation index to cybermetrics*. Lanham: Scarecrow.
Diamond, A. (1984). An economic model of the life-cycle research productivity of scientists. *Scientometrics, 6*(3), 189–196.
Diamond, A. (1986). What is a citation worth? *The Journal of Human Resources, 21*(2), 200–215.
Diamond, N., & Graham, H. (2000). How should we rate research universities? *Change, 32*(4), 21–33.
Dolan, R., & Schmidt, R. (1994). Modeling institutional production of higher education. *Economics of Education Review, 13*(3), 197–213.
Dundar, H., & Lewis, D. (1995). Departmental productivity in American universities: Economies of scale and scope. *Economics of Education Review, 14*(2), 119–144.
Etzkowitz, H. (2008). *The triple helix: University-industry-government innovation in action*. New York: Routledge.
Etzkowitz, H., & Dzisah, J. (2008). Rethinking development: Circulation in the triple helix. *Technology Analysis & Strategic Management, 20*(6), 653–666.
Garfield, E. (1979). *Citation indexing: Its theory and applications in science, technology, and humanities*. New York: Free Press.
Geiger, R. L. (2004). *Research & relevant knowledge: American universities since WWII*. New Brunswick: Transaction Press.
Geiger, R., & Sa, C. (2008). *Tapping the riches of science: Universities and the promise of economic growth*. Boston: Harvard University Press.
Gerrity, D., & McKenzie, R. (1978). The ranking of southern economics departments: New criterion and further evidence. *Southern Economic Journal, 45*(2), 608–614.
Huang, M., & Chang, Y. (2008). Characteristics of research output in social sciences and humanities: From a research evaluation perspective. *Journal of the American Society for Information Science and Technology, 59*(11), 1819–1828.
James, A. D. (2009). Reevaluating the role of defense and security R&D in the innovative system: Reform of the UK government defence research establishments. *Journal of Technology Transfer, 34*(5), 505–523.
Johnes, J., Taylor, J., & Francis, B. (1993). The research performance of UK universities: A statistical analysis of the results of the 1989 research selectivity exercise. *Journal of the Royal Statistical Society. Series A, 156*(2), 271–286.
Krampen, G., Becker, R., Wahner, U., & Montada, L. (2007). On the validity of citation counting in science evaluation: Content analyses of references and citations in psychological publications. *Scientometrics, 71*(2), 191–202.

Laband, D. (1985). An evaluation of 50 'ranked' economics departments by quantity and quality of faculty *publications* and graduate student placement and research success. *Southern Economic Journal, 52*, 216–240.
Libaers, D. (2009). Reevaluating the role of defense and security R&D in the innovation system: industry relationships of DoD-funded academics and institutional changes in US university systems. *Journal of Technology Transfer, 34*(5), 474–489.
Lindsey, D. (1989). Using citation counts as a measure of quality in science. *Scientometrics, 15*(3–4), 189–203.
Long, J., & McGinnis, R. (1982). On adjusting productivity measures for multiple authorship. *Scientometrics, 4*(5), 379–387.
Maher, B. (1996). The NRC's report on research-doctorate programs: It's uses and misuses. *Change, 28*(6), 54–59.
Marsh, H. W., & Hattie, J. (2002). The relation between research productivity and teaching effectiveness. *The Journal of Higher Education, 73*(5), 603–641.
McCormick, A., & Zhao, C. (2005). Rethinking and reframing the Carnegie classification. *Change, 37*(5), 51–57.
Moed, H. (2005). *Citation analysis in research evaluation*. New York: Springer.
Moed, H. (2006). *Bibliometric rankings of world universities* (CWTS Report 2006-01). Leiden: Centre for Science and Technology Studies (CWTS).
Moed, H., Burger, W., Frankfort, J., & Van Raan, A. (1985). The use of bibliometric data for the measurement of university research performance. *Research Policy, 14*(3), 131–149.
National Center for Education Statistics. (2009). *Digest of education statistics 2008*. Washington DC: Institute of Education Sciences.
Nederhof, A. (2006). Bibliometric monitoring of research performance in the social sciences and the humanities: A review. *Scientometrics, 66*(1), 81–100.
Porta, M., Fernandez, E., & Puigdomenech, E. (2006). Book citations: Influence of epidemiologic thought in the academic community. *Revista de Saúde Pública, 40*, 50–56.
Porter, S., & Toutkoushian, R. (2006). Institutional research productivity and the connection to average student quality and overall reputation. *Economics of Education Review, 25*(6), 605–617.
Roy, R., Roy, N., & Johnson, G. (1983). Approximating total citation counts from first author counts and from total papers. *Scientometrics, 5*(2), 117–124.
Seglen, P. O. (1997). Why the impact factor of journals should not be used for evaluating research. *British Medical Journal, 314*(7079), 498–502.
Seol, S., & Park, J. (2008). Knowledge sources of innovation studies in Korea: A citation analysis. *Scientometrics, 75*(1), 3–20.
Shin, J. (2009). Building world-class research university: The Brain Korea 21 project. *Higher Education, 58*(5), 669–688.
Shin, K., & Putnam, R. (1982). Age and academic – professional honors. *Journal of Gerontology, 37*(2), 220–229.
Slaughter, S., & Rhoades, G. (2004). *Academic capitalism and the new economy*. Baltimore: Johns Hopkins.
Stern, M. (1983). Characteristics of the literature and intellectual acceptance of scholarly monographs. *College & Research Libraries, 44*(4), 199–209.
Tan, D. (1986). The assessment of quality in higher education: A critical review of the literature and research. *Research in Higher Education, 24*(3), 223–265.
Tang, R. (2008). Citation characteristics and intellectual acceptance of scholarly monographs. *College & Research Libraries, 69*(4), 356–369.
Toutkoushian, R. (1994). Using citations to measure sex discrimination in faculty salaries. *The Review of Higher Education, 18*(1), 61–82.
Toutkoushian, R., Porter, S., Danielson, C., & Hollis, P. (2003). Using publication counts to measure an institution's research productivity. *Research in Higher Education, 44*(2), 121–148.
University System of Georgia Annual Financial Report FY 2007. Retrieved on October 18, 2009, from http://www.usg.edu/fiscal_affairs/reporting/annual_fin_rep/2007/afr2007_cons_pdf.pdf

Wardle, D. (1995). Journal citation impact factors and parochial citation practices. *Bulletin of the Ecological Society of America, 76*, 102–104.

Webster, D., & Skinner, T. (1996). Rating PhD programs: What the NRC report says… and doesn't say. *Change, 28*(3), 22–32, 34–44.

Yates, S., & Chapman, K. (2005). An examination of the use of monographs in the communication journal literature. *Behavioral and Social Sciences Librarian, 26*(11), 39–51.

Zainab, A., & Goi, S. (1997). Characteristics of citations used by humanities researchers. *Malaysian Journal of Library & Information Science, 2*(2), 19–36.

Chapter 8
Peer Review and Bibliometric: Potentials and Problems

Lutz Bornmann

8.1 Introduction

Among the various human activities, activities in science are those that are the most subject to evaluation by peers (Laloë and Mosseri 2009). Such evaluations determine, among ranking positions of universities, who gets which job, who gets tenure, and who gets which awards and honors (Feist 2006). For the THE – QS World University Rankings, the assessment by peers is the centerpiece of the ranking process; peer review is also a major indicator in the US News & World Report rankings (Enserink 2007). "By defining losers and winners in the competition for positions, grants, publication of results, and all kinds of awards, peer review is a central social control institution in the research community" (Langfeldt 2006: 32). Research evaluation systems in the various countries of the world (e.g., the British research assessment exercise) are normally based on peer review. The edited book of Whitley and Gläser (2007) shows how these systems are changing the organization of scientific knowledge production and universities in the countries involved (Moed 2008).

Aside from the selection of manuscripts for publication in journals, the most common contemporary application of peer review in scientific research is for the selection of fellowship and grant applications. Peers or colleagues, asked to evaluate applications or manuscripts in a peer review process, take on the responsibility for assuring high standards in various research disciplines. Although peers active in the same field might be blind-sided by adherence to the same specialist group, they "are said to be in the best position to know whether quality standards have been met and a contribution to knowledge made" (Eisenhart 2002: 241). Peer evaluation in

L. Bornmann (✉)
Max Planck Society, Hofgartenstr. 8, D-80539 Munich, Germany
e-mail: bornmann@gv.mpg.de

J.C. Shin et al. (eds.), *University Rankings*, The Changing Academy – The Changing Academic Profession in International Comparative Perspective 3,
DOI 10.1007/978-94-007-1116-7_8,

research thus entails a process by which a selective jury of equals, active in a given scientific field, convenes to evaluate the undertaking of scientific activity or its outcomes. Such a jury of equals may be consulted as a group or individually, without the need for personal contacts among the evaluators. The peer review process lets the active producers of science, the experts, become the "gatekeepers" of science (McClellan 2003).

Proponents of the peer review system argue that it is more effective than any other known instrument for self-regulation in science. Putting it into a wider context, according to the critical rationalism of Popper (1961) intellectual life and institutions should be arranged to provide "maximum criticism, in order to counteract and eliminate as much intellectual error as possible" (Bartley 1984: 113). Evidence supports the view that peer review improves the quality of the reporting of research results (Goodman et al. 1994; Pierie et al. 1996). As a proponent of peer review, Abelson writes (1980): "The most important and effective mechanism for attaining good standards of quality in journals is the peer review system" (p. 62). According to Shatz (2004) journal peer review "motivates scholars to produce their best, provides feedback that substantially improves work which is submitted, and enables scholars to identify products they will find worth reading" (p. 30).

Critics of peer review argue that (1) reviewers rarely agree on whether or not to recommend that a manuscript be published or a research grant be awarded, thus making for poor *reliability* of the peer review process; (2) reviewers' recommendations are frequently biased, that is, judgments are not based solely on scientific merit, but are also influenced by personal attributes of the authors, applicants, or the reviewers themselves (where the *fairness* of the process is not given); and (3) the process lacks *predictive validity*, since there is little or no relationship between the reviewers' judgments and the subsequent usefulness of the work to the scientific community, as indicated by the frequency of citations of the work in later scientific papers. According to Butler (2007), the assessment by peers as an indicator in the US News & World Report university ranking implies a false precision and authority. For further criticisms on scientific peer review see Hames (2007) and Schmelkin (2006).

In recent years, a number of published studies have addressed these criticisms raised about scientific peer review. From the beginning, this research on peer review has focused on the evaluation of manuscripts and (fellowship or grant) applications.

"The peer review process that scholarly publications undergo may be interpreted as a sign of 'quality.' But to many, a publication constitutes nothing more than an 'offer' to the scientific community. It is the subsequent reception of that offer that certifies the actual 'impact' of a publication" (Schneider 2009: 366). Formal citations are meant to show that a publication has made use of the contents of other publications (research results, others' ideas, and so on). Citation counts (the number of citations) are used in research evaluation as an indicator of the impact of the research: "The *impact* of a piece of research is the degree to which it has been useful to other researchers" (Shadbolt et al. 2006: 202). According to the Research Evaluation and Policy Project (2005), there is an emerging trend to regard impact, the measurable part of quality, as a proxy measure for quality in total. For Lindsey,

citations are "our most *reliable* convenient measure of quality in science – a measure that will continue to be widely used" (Lindsey 1989: 201).

In research evaluation, citation analyses have been conducted for assessment of national science policies and disciplinary development (e.g., Lewison 1998; Oppenheim 1995, 1997; Tijssen et al. 2002), departments and research laboratories (e.g., Bayer and Folger 1966; Narin 1976), books and journals (e.g., Garfield 1972; Nicolaisen 2002), and individual scientists (e.g., Cole and Cole 1973; Garfield 1970). Besides peer review with a 40% weighting, the THE – QS World University Rankings gives the indicator "citations per faculty" a 20% weighting. The Leiden Ranking system is entirely based on bibliometric indicators (Enserink 2007).

Citation counts are attractive raw data for the evaluation of research output. Because they are "unobtrusive measures that do not require the cooperation of a respondent and do not themselves contaminate the response (i.e., they are non-reactive)" (Smith 1981: 84), citation rates are seen as an *objective* quantitative indicator for scientific success and are held to be a valuable complement to qualitative methods for research evaluation, such as peer review (Daniel 2005; Garfield and Welljamsdorof 1992). Scientific "reward came primarily in the form of recognition rather than money, an insight that helps account for the importance scientists place upon citation as a reward system … This idea of citation as a kind of stand-in for direct economic reward – what is sometimes called the citation credit cycle – is often seen as a feature of academic reward generally" (Kellogg 2006: 3).

However, out in the early 1970s, Eugene Garfield, the founder of the Institute of Scientific Information (ISI, now Thomson Reuters, Philadelphia, PA, USA) pointed out that citation counts are a function of many variables besides scientific quality (Garfield 1972). In a recently published paper, Laloë and Mosseri (2009) state that bibliometric methods "do contain information about scientific quality, but this 'signal' is buried in a 'noise' created by a dependence on many other variables" (p. 27). Up to now, a number of variables that generally influence citation counts have emerged in bibliometric studies. Lawani (1986) and other researchers established, for example, that there is a positive relation between the number of co-authors of a publication and its citation counts; a higher number of co-authors is usually associated with a higher number of citations. Based on the findings of these studies, the number of co-authors and other general influencing factors should be taken into consideration in evaluative bibliometric studies.

Since research evaluation is an area of increasing importance, it is necessary that the application of peer review and impact measures (citation counts) is done well and professionally (see here de Vries et al. 2009). For that, background information about empirical findings on both evaluation instruments is necessary (especially findings that are related to their problems). In Sect. 8.2 of this chapter, an overview is provided on studies that have conducted meta-evaluations of peer review procedures, because a literature search found no empirical studies on peer review in the context of university rankings, Sect. 8.2 focuses on journal, fellowship, and grant peer review. In general, the results are applicable to the use of peer review in the context of university rankings. Sect. 8.3 gives an overview on studies that have investigated citation counts to identify general influencing factors.

8.2 Research on Journal, Fellowship, and Grant Peer Review

8.2.1 Agreement Among Reviewers (Reliability)

"In everyday life, intersubjectivity is equated with realism" (Ziman 2000: 106). The scientific discourse is also distinguished by a striving for consensus. Scientific activity would clearly be impossible unless scientists could come to similar conclusions. According to Wiley (2008) "just as results from lab experiments provide clues to an underlying biological process, reviewer comments are also clues to an underlying reality (they did not like your grant for some reason). For example, if all reviewers mention the same point, then it is a good bet that it is important and real." An established consensus among scientists must of course be a voluntary one achieved under conditions of free and open criticism (Ziman 2000). The norms of the ethos of science make these conditions possible and regulate them (Merton 1942): The norms of communalism (scientific knowledge should be made public knowledge) and universalism (knowledge claims should be judged impersonally, independently of their source) envisage eventual agreement. "But the norm of 'organized skepticism', which energizes critical debates, rules out any official procedure for closing them. Consensus and dissensus are thus promoted simultaneously" (Ziman 2000: 255) by the norms of the ethos of science.

If a submission (manuscript or application) meets scientific standards and contributes to the advancement of science, one would expect that two or more reviewers will agree on its value. This, however, is frequently not the case. Ernst et al. (1993) offer a dramatic demonstration of the unreliability of the journal peer review process. Copies of one paper submitted to a medical journal were sent simultaneously to 45 experts. They were asked to express their opinion of the paper with the journal's standard questionnaire judging eight quality criteria on a numerical scale from 5 (excellent) to 1 (unacceptable). The 31 correctly filled forms demonstrated poor reliability with extreme judgments ranging from "unacceptable" to "excellent" for most criteria. The results of studies on reliability in journal peer review indicate that the levels of inter-reviewer agreement, when corrected for chance, generally fall in the range from 0.20 to 0.40 (Bornmann 2011), which indicates a relatively low level of reviewer agreement.

Reviewer disagreement is not always seen as a negative factor however, as many see it as a positive method of evaluating a manuscript from a number of different perspectives. If reviewers are selected for their opposing viewpoints or expertise, a high degree of reviewer agreement should not be expected. It can even be argued that too much agreement is in fact a sign that the review process is not working well, that reviewers are not properly selected for diversity, and that some are redundant. Whether the comments of reviewers are in fact based on different perspectives is a question that has been examined by only a few empirical studies (Weller 2002). One study, for example, showed that reviewers of the same manuscript simply commented on different aspects of the manuscript: "In the typical case, two reviews of the same paper had no critical point in common … [T]hey wrote about different

topics, each making points that were appropriate and accurate. As a consequence, their recommendations about editorial decisions showed hardly any agreement" (Fiske and Fogg 1990: 591).

The fate of a manuscript depends on which small sample of reviewers influences the editorial decision, as research such as that of Bornmann and Daniel (2009a, 2010) for the *Angewandte Chemie International Edition* (AC-IE) indicates. In AC-IE's peer review process, a manuscript is generally published only if two reviewers rate the results of the study as important and also recommend publication in the journal (what the editors have called the "clear-cut" rule). Even though the "clear-cut" rule is based on two reviewer reports, submitted manuscripts generally go out to three reviewers in total. An editor explains this process in a letter to an author as follows: "Many papers are sent initially to three referees (as in this case), but in today's increasingly busy climate there are many referees unable to review papers because of other commitments. On the other hand, we have a responsibility to authors to make a rapid and fair decision on the outcome of papers." For 23% of those manuscripts, for which a third reviewer report arrived after the editorial decision was made (37 of 162), this rule would have led to a different decision if the third report had replaced either of the others. Consequently, even if the editor considered all three reviewers to be suitable to review a manuscript, the editor would have needed to make a different decision based on the changed situation.

8.2.2 Fairness of the Peer Review Process

According to Merton (1942) the functional goal of science is the expansion of potentially true and secure knowledge. To fulfill this function in society, the ethos of science was developed. The norm of universalism prescribes that the evaluation of scientific contributions should be based upon objective scientific criteria. Journal submissions or grant applications are not supposed to be judged according to the attributes of the author/applicant or the personal biases of the reviewer, editor, or program manager (Ziman 2000). "First, universalism requires that when a scientist offers a contribution to scientific knowledge, the community's assessment of the validity of that claim should not be influenced by personal or social attributes of the scientist …Second, universalism requires that a scientist be fairly rewarded for contributions to the body of scientific knowledge …Particularism, in contrast, involves the use of functionally irrelevant characteristics, such as sex and race, as a basis for making claims and gaining rewards in science" (Long and Fox 1995: 46). To the degree that particularism influences how claims are made and rewards are gained, the fairness of the peer review process is at risk (Godlee and Dickersin 2003).

Ever since Kuhn (1962) discussed the significance of different scientific views or paradigmatic views for the evaluation of scientific contributions in his seminal work *The structure of scientific revolutions* (see here also Mallard et al. 2009), researchers have expressed increasing doubt about the norm-ruled objective evaluation of scientific work (Hemlin 1996). Above all, proponents of social

constructivism have expressed such doubts since the 1970s. For Cole (1992) the research of the constructivists supports a new view of science which casts doubt on the existence of a set of rational criteria. The most valuable of insights into scientists' actions, social constructivist research, according to Sismondo (1993), has brought about the recognition that "social objects in science exist and act as causes of, and constraints on, scientists' actions" (p. 548). Because reviewers are human, factors which cannot be predicted, controlled, or standardized influence their writing of reviews, according to Shashok (2005).

Reviews of peer review research (Hojat et al. 2003; Owen 1982; Pruthi et al. 1997; Ross 1980; Sharp 1990; Wood and Wessely 2003) name up to 25 potential sources of bias in peer review. In these studies, it is usual to call any feature of an assessor's cognitive or attitudinal mind-set that could interfere with an objective judgment, a bias (Shatz 2004). Factors that appear to bias assessors' objective judgments with respect to a manuscript or an application include nationality, gender of the author or applicant, and the area of research from which the work originates. Other studies show that replication studies and research that lead to statistically insignificant findings stand a rather low chance of being judged favorably by peer reviewers.

Research on bias in peer review faces two serious problems. First, the research findings on bias are inconsistent. For example, some studies investigating gender bias in journal review processes point out that women scientists are at a disadvantage. However, a similar number of studies report no gender effects or mixed results. Second, it is almost impossible to establish unambiguously whether work from a particular group of scientists (e.g., junior or senior scientists) receives better reviews and thus a higher acceptance rate due to preferential biases affecting the review and decision-making process, or if favorable review and favorable judgments in peer review are simply a consequence of the high scientific quality of the corresponding manuscripts or applications.

Presumably, it will never be possible to eliminate all doubts regarding the fairness of the review process. Because reviewers are human, their behavior – whether performing their salaried duties, enjoying their leisure time, or writing reviews – is influenced by factors that cannot be predicted, controlled or standardized (Shashok 2005). Therefore, it is important that the peer review process should be further studied. Any evidence of bias in judgments should be uncovered for purposes of correction and modification of the process (Geisler 2001; Godlee and Dickersin 2003).

8.2.3 Predictive Validity of the Peer Review Process

The goal for peer review of grant/ fellowship applications and manuscripts is usually to select the "best" from among the work submitted (Smith 2006). In investigating the predictive validity of the peer review process, the question arises as to whether this goal is actually achieved, that is, whether indeed the "best" applications or manuscripts are funded or published. The validity of judgments in peer review is often questioned. For example, the former editor of the journal *Lancet*,

Sir Theodore Fox (1965), writes on the validity of editorial decisions: "When I divide the week's contributions into two piles – one that we are going to publish and the other that we are going to return – I wonder whether it would make any real difference to the journal or its readers if I exchanged one pile for another" (p. 8). The selection function is considered to be a difficult research topic to investigate. According to Jayasinghe et al. (2001) and Figueredo (2006), there exists no mathematical formula or uniform definition as to what makes a manuscript "worthy of publication," or what makes a research proposal "worthy of funding" (see also Smith 2006).

For the investigation of the predictive validity of the peer review process, the impact of papers accepted or rejected (but published elsewhere) in peer reviewed journals, or the impact of papers that were published by applicants whose proposals were either accepted or rejected in grant or fellowship peer reviews, are compared. Because the number of citations of a publication reflects its international impact (Borgman and Furner 2002; Nicolaisen 2007) and because of the lack of other operationalizable indicators, it is a common approach in peer review research to evaluate the success of the process on the basis of citation counts (see Sect. 8.3). Scientific judgments on submissions (manuscripts or applications) are said to show predictive validity in peer review research, if the citation counts of manuscripts accepted for publication (or manuscripts published by accepted applicants) and manuscripts rejected by a journal but then published elsewhere (or manuscripts published by rejected applicants) differ statistically significantly.

Up until now, only a few studies have conducted analyses which examine citation counts from individual papers as the basis for assessing predictive validity in peer reviews. A literature research found only six empirical studies on the level of predictive validity associated with the journal peer review process. Research in this area is extremely labor-intensive, since a validity test requires information and citation counts regarding the fate of rejected manuscripts (Bornstein 1991). The editor of the *Journal of Clinical Investigation* (Wilson 1978) has undertaken his own investigation into the question of predictive validity. Daniel (1993) and Bornmann and Daniel (2008a, b) investigated the peer review process of AC-IE, and Opthof et al. (2000) did the same for *Cardiovascular Research*. McDonald et al. (2009) and Bornmann et al. (2010) examined the predictive validity of the editorial decisions for the *American Journal of Neuroradiology* and *Atmospheric Chemistry and Physics*. All six studies confirmed that the editorial decisions (acceptance or rejection) for the various journals appear to reflect a rather high degree of predictive validity, if citation counts are employed as validity criteria.

Eight studies on the assessment of citation counts, as a basis of predictive validity in selection decisions in fellowship or grant peer reviews, have been published in recent years according to a literature search. The studies by Armstrong et al. (1997) on the Heart and Stroke Foundation of Canada (HSFC, Ottawa), the studies by Bornmann and Daniel (2005b, 2006) on the Boehringer Ingelheim Fonds (Heidesheim, Germany), and by Bornmann et al. (2008) on the European Molecular Biology Organization (Heidelberg, Germany), and the study of Reinhart (2009) on the Swiss National Science Foundation (Bern) confirm the predictive validity of the

selection decisions, whereas the studies by Hornbostel et al. (2009) on the Emmy Noether Programme of the German Research Foundation (Bonn) and by Melin and Danell (2006) on the Swedish Foundation for Strategic Research (Stockholm) showed no significant differences between the performance of accepted and rejected applicants. Van den Besselaar and Leydesdorff (2007) report on contradictory results regarding the Council for Social Scientific Research of the Netherlands Organization for Scientific Research (Den Haag). The study by Carter (1982) investigated the association between (1) assessments given by the reviewers for the National Institutes of Health (Bethesda, MD, USA) regarding applicants for research funding, and (2) the number of citations, which articles in journals produced under the grants have obtained. This study showed that better votes in fact correlate with more frequent citations; however, the correlation coefficient was low.

Unlike the clearer results for journal peer reviews, contradictory results emerge in research on fellowship or grant peer reviews. Some studies confirm the predictive validity of peer reviews, while the results of other studies leave room for doubt about their predictive validity.

8.3 Research on Citation Counts as Bibliometric Indicator

The research activity of a group of scientists, publication of their findings, and citation of the publications by colleagues in the field are all social activities. This means that citation counts for the group's publications are not only an indicator of the impact of their scientific work on the advancement of scientific knowledge (as stated by the *normative theory* of citing; see a description of the theories of citing in the next section). According to the *social constructivist view* on citing, citations also reflect (social) factors that do not have to do with the accepted conventions of scholarly publishing (Bornmann and Daniel 2008c). "There are 'imperfections' in the scientific communications system, the result of which is that the importance of a paper may not be identical with its impact. The 'impact' of a publication describes its actual influence on surrounding research activities at a given time. While this will depend partly on its importance, it may also be affected by such factors as the location of the author, and the prestige, language, and availability of the publishing journal" (Martin and Irvine 1983: 70). Bibliometric studies published in recent years have revealed the general influence of this and a number of other factors on citation counts (Peters and van Raan 1994).

8.3.1 Theoretical Approaches to Explaining Citing

Two competing theories of citing have been developed in past decades, both of them situated within broader social theories of science. One is often denoted as the normative theory of citing and the other as the social constructivist view of citing.

The *normative theory*, following Robert K. Merton's sociological theory of science (Merton 1973), basically states that scientists give credit to colleagues whose work they use by citing that work. Thus, citations represent intellectual or cognitive influence on scientific work. Merton (1988) expressed this aspect as follows: "The reference serves both instrumental and symbolic functions in the transmission and enlargement of knowledge. Instrumentally, it tells us of work we may not have known before, some of which may hold further interest for us; symbolically, it registers in the enduring archives the intellectual property of the acknowledged source by providing a pellet of peer recognition of the knowledge claim, accepted or expressly rejected, that was made in that source" (p. 622, see also Merton 1957; Merton 1968).

The *social constructivist view* on citing is grounded in the constructivist sociology of science (see, e.g., Collins 2004; Knorr-Cetina 1981; Latour and Woolgar 1979). This view casts doubt on the assumptions of normative theory and questions the validity of evaluative citation analysis. Constructivists argue that the cognitive content of articles has little influence on how they are received. Scientific knowledge is socially constructed through the manipulation of political and financial resources, and the use of rhetorical devices (Knorr-Cetina 1991). For this reason, citations cannot be satisfactorily described unidimensionally through the intellectual content of the article itself. The probability of being cited depends on many factors that are not related to the accepted conventions of scholarly publishing. In the next section, an overview of these factors is given.

8.3.2 Factors that Influence Citation Counts in General

8.3.2.1 Time-Dependent Factors

Due to the exponential increase in scientific output, citations become more probable from year to year. Beyond that, it has been shown that the more frequently a publication is cited, the more frequently it will be cited in future; in other words, the expected number of future citations is a linear function of the current number. Cozzens (1985) calls this phenomenon "success-breeds-success," and it holds true not only for highly-cited publications, but also for highly-cited scientists (Garfield 2002). However, according to Jensen et al. (2009) "the assumption of a constant citation rate *unlimited in time* is not supported by bibliometric data" (p. 474).

8.3.2.2 Field-Dependent Factors

Citation practices vary between science and social science fields (Castellano and Radicchi 2009; Hurt 1987; Radicchi et al. 2008) and even within different areas (or clusters) within a single subfield (Bornmann and Daniel 2009b). In some fields, researchers cite recent literature more frequently than in others. As the chance of

being cited is related to the number of publications in the field, small fields attract far fewer citations than more general fields (King 1987).

8.3.2.3 Journal-Dependent Factors

Ayres and Vars (2000) found that the first article in the journal tended to produce more citations than the later ones, perhaps because the editors recognized such articles to be especially important. Stewart (1983) argued that the citation of an article may depend on the frequency of publication of journals containing related articles. Furthermore, journal accessibility, visibility, and internationality as well as the impact, quality, or prestige of the journal may influence the probability of citations (Judge et al. 2007; Larivière and Gingras 2010; Leimu and Koricheva 2005).

8.3.2.4 Article-Dependent Factors

Citation characteristics of methodology articles, review articles, research articles, letters, and notes as well as articles, chapters, and books differ considerably (Lundberg 2007). There is also a positive correlation between the citation frequency of publications and (1) the number of co-authors of the work (Lansingh and Carter 2009), and (2) the number (Fok and Franses 2007) and the impact (Boyack and Klavans 2005) of the references within the work. Moreover, as longer articles have more content that can be cited than do shorter articles, the sheer size of an article influences whether it is cited (Hudson 2007).

8.3.2.5 Author- /Reader-Dependent Factors

The language a paper is written in (Kellsey and Knievel 2004; Lawani 1977) and cultural barriers (Carpenter and Narin 1981; Menou 1983) influence the probability of citations. Results from Mählck and Persson (2000), White (2001), and Sandström et al. (2005) show that citations are affected by social networks, and that authors cite primarily works by authors with whom they are personally acquainted. Cronin (2005) finds this hardly surprising, as it is to be expected that personal ties become manifest and strengthened, resulting in greater reciprocal exchange of citations over time.

8.3.2.6 Literature- and Citation Database–Dependent Factors

Free online availability of publications influences the probability of citations (Lawrence 2001; McDonald 2007). Citation analyses cannot be any more accurate than the raw material used (Smith 1981; van Raan 2005b). The incorrect citing of sources is unfortunately far from uncommon. Evans et al. (1990) checked the references in papers in three medical journals and determined that 48% were

incorrect: "The data support the hypothesis that authors do not check their references or may not even read them" (p. 1353). In a similar investigation, Eichorn and Yankauer (1987) found that "thirty-one percent of the 150 references had citation errors, one out of 10 being a major error (reference not locatable)" (p. 1011). Unver et al. (2009) found errors in references "in about 30% of current physical therapy and rehabilitation articles" (p. 744). Furthermore, the data in the literature data bases like Web of Science (WoS, Thomson Reuters) or Scopus (Elsevier) are not "homogeneous, since the entry of data has fluctuated in time with the persons in charge of it. It, therefore, requires a specialist to make the necessary series of corrections" (Laloë and Mosseri 2009: 28). Finally, according to Butler (2007) "Thomson Scientific's [now Thomson Reuters] ISI citation data are notoriously poor for use in rankings; names of institutions are spelled differently from one article to the next, and university affiliations are sometimes omitted altogether. After cleaning up ISI data on all UK papers for such effects, the Leeds-based consultancy Evidence Ltd. found the true number of papers from the University of Oxford, for example, to be 40% higher than listed by ISI, says director Jonathan Adams" (p. 514, see also Bar-Ilan 2009). Errors in these data are especially serious, as most of the rankings are based on Thomson Reuter's data (Buela-Casal et al. 2007).

8.4 Discussions

Buela-Casal et al. (2007) presented a comparative study of four well-known international university rankings. Their results show that generally peer review and citation counts play an important role as indicators in these rankings. Although university rankings are a growing phenomenon in higher education worldwide (Merisotis and Sadlak 2005), there is surprisingly little empirical research on the use of these dominating indicators. The research on peer review and citation counts (still) refers to other areas. However, as the results of this research are generalizable, this chapter has provided a research overview including the most important studies.

Against the backdrop of these studies, it can be assumed that peer assessments given for rankings are affected by disagreements among independent peers as well as biases and a lack of predictive validity: (1) One and the same university will be assessed differently by independent peers; (2) other criteria than scientific quality will influence the universities' assessments; (3) the assessments might not be correlated with other indicators of scientific quality. Referring to citation counts, the research points out that this impact measure is affected by some general influencing factors. Thus, citation counts only measure an aspect of the scientific quality of universities. In the following paragraphs, we will summarize and discuss the most important findings presented in Sects. 8.2 and 8.3.

In recent years, a number of published studies have taken up and investigated the criticisms that have been raised against the scientific peer review process. Some important studies were presented in Sect. 8.2. To recapitulate the study results

published so far on the *reliability* of peer review: Most studies report a low level of agreement between reviewers' judgments. However, very few studies have investigated reviewer agreement with the purpose of identifying the actual reasons behind reviewer disagreement (e.g., by carrying out comparative content analyses of reviewers' comment sheets). LaFollette (1992), for example, noted the scarcity of research on such questions as how reviewers apply standards and the specific criteria established for making a decision. In-depth studies that address these issues might prove to be fruitful avenues for future investigation (Weller 2002). This research should primarily dedicate itself to the dislocational component in the judgment of reviewers as well as differences in strictness or leniency in reviewers' judgments (Eckes 2004; Lienert 1987).

Although reviewers like to believe that they choose the "best" based on objective criteria, "decisions are influenced by factors – including biases about race, sex, geographic location of a university, and age – that have nothing to do with the quality of the person or work being evaluated" (National Academy of Sciences 2006). Considering that peers are not prophets but ordinary human beings with their own opinions, strengths, and weaknesses (Ehses 2004), a number of studies have already worked on *potential sources of bias* in peer review. Although numerous studies have shown an association between potential sources of bias and judgments in peer review and thus called into question the fairness of the process itself, the research on these biases faces two fundamental problems that make generalization of the findings difficult. On the one hand, the various studies have yielded quite heterogeneous results. Some studies have proven the indisputable effects of potential sources of bias; in other studies, they showed moderate or slight effects. A second principal problem that affects bias research in general is the pervasive lack of experimental studies. This shortage makes it impossible to establish unambiguously whether work from a particular group of scientists receives better reviews due to biases in the review and decision-making process, or if favorable reviews and greater success in the selection process are simply a consequence of the scientific merit of the corresponding group of proposals or manuscripts.

The few studies, which have examined the *predictive validity* of journal peer review on the basis of citation counts, confirm that a peer review represents a quality filter and works as an instrument for the self-regulation of science. Concerning fellowship or grant peer reviews, there are more studies which have investigated the predictive validity of selection decisions on the basis of citation counts. Compared with journal peer reviews, these studies have provided heterogeneous results; some studies can confirm the predictive validity of peer reviews, whereas the results of other studies leave that in doubt.

The heterogeneous results on fellowship and grant peer review can be attributed to the fact that "funding decisions are inherently speculative because the work has not yet been done" (Stamps 1997: 4). Whereas in a journal peer review the *results* of the research are assessed, a grant and fellowship peer review is principally an evaluation of the *potential* of the proposed research (Bornmann and Daniel 2005a). Evaluating the application involves deciding whether the proposed research is significant, determining whether the specific plans for investigation are feasible, and

evaluating the competence of the applicant (Cole 1992). Fellowship or grant peer reviews – when compared to journal peer reviews – are perceived as entailing a heightened risk for judgments and decisions with low predictive validity. Accordingly, it is expected that studies on grant or fellowship peer reviews are less likely than studies on journal peer reviews to be able to confirm the predictive validity.

In recent years, besides the qualitative form of research evaluation, the peer review system, the quantitative form has become more and more important. "Measurement of research excellence and quality is an issue that has increasingly interested governments, universities, and funding bodies as measures of accountability and quality are sought" (Steele et al. 2006: 278). Weingart (2005a) notes that a really enthusiastic acceptance of bibliometric figures for evaluative purposes or for comparing the research success of scientists can be observed today. University rankings are normally based on bibliometric measures. The United Kingdom is planning to allocate government funding for research by universities in large part using bibliometric indicators: "The Government has a firm presumption that after the 2008 RAE [Research Assessment Exercise], the system for assessing research quality and allocating 'quality-related' (QR) research funding to universities from the Department for Education and Skills will be mainly metrics-based" (UK Office of Science and Technology 2006: 3). With the easy availability of bibliometric data and ready-to-use tools for generating bibliometric indicators for evaluation purposes, there is a danger of improper use.

As noted above, two competing theories of citing were developed in past decades: the normative theory of citing and the social constructive approach to citing. Following normative theory, the reasons why scientists cite documents are that the documents are relevant to their topic and provide useful background for their research and in order to acknowledge intellectual debt. The social constructive view on citing contradicts these assumptions. According to this view, citations are a social psychological process, not free of personal bias or social pressures and probably not made for the same reasons. While Cronin (1984) finds the existence of two competing theories of citing behavior hardly surprising, as the construction of scientific theory is generally characterized by ambivalence, for Liu (1997) and Weingart (2005b), the long-term oversimplification of thinking in terms of two theories reflects the absence of one satisfactory and accepted theory on which the better informed use of citation indicators could be based. Whereas Liu (1997) and Nicolaisen (2003) see the dynamic linkage of both theories as a necessary step in the quest for a satisfactory theory of citation, Garfield (1998) states: "There is no way to predict whether a particular citation (use of a reference by a new author) will be 'relevant'" (p. 70).

The results of the studies presented in Sect. 8.3 suggest that not only the content of scientific work, but also other, in part non-scientific, factors play a role in citing. Citations can therefore be viewed as a complex, multi-dimensional and not a unidimensional phenomenon. The reasons authors cite can vary from scientist to scientist. On the basis of the available findings, should we then conclude that citation counts are not appropriate indicators of the impact of research? Are citation counts

not suitable for use in university rankings? Not so, says van Raan (2005a): "So undoubtedly the process of citation is a complex one, and it certainly not provides an 'ideal' monitor on scientific performance. This is particularly the case at a statistically low aggregation level, e.g., the individual researcher. There is, however, sufficient evidence that these reference motives are not so different or 'randomly given' to such an extent that the phenomenon of citation would lose its role as a reliable measure of impact. Therefore, application of citation analysis to the entire work, the 'oeuvre' of *a group of researchers as a whole over a longer period of time*, does yield in many situations a strong indicator of scientific performance" (p. 134–135, see also Laloë and Mosseri 2009).

Research on the predictive validity of peer review indicates that peer review is generally a credible method for evaluation of manuscripts and – in part – of grant and fellowship applications. But this overview of the reliability and fairness of the peer review process shows that there are also problems with peer reviews. However, despite its flaws, having scientists judge each other's work is widely considered to be the "least bad way" to weed out weak work (Enserink 2001). In a similar manner, bibliometric indicators do have specific drawbacks. However, on a higher aggregation level (a larger group of scientists), it seems to be a reliable indicator of research impact. It has been frequently recommended that peer review should be used for the evaluation of scientific work and should be supplemented with bibliometrics (and other metrics of science) to yield a broader and powerful methodology for assessment of scientific advancement (Geisler 2001; van Raan 1996). Thus, the combination of both indicators in university rankings seems to be a sensible way to build on the strengths and compensate for the weaknesses of both evaluative instruments.

References

Abelson, P. H. (1980). Scientific communication. *Science, 209*(4452), 60–62.

Armstrong, P. W., Caverson, M. M., Adams, L., Taylor, M., & Olley, P. M. (1997). Evaluation of the heart and stroke foundation of Canada research Scholarship Program: Research productivity and impact. *The Canadian Journal of Cardiology, 13*(5), 507–516.

Ayres, I., & Vars, F. E. (2000). Determinants of citations to articles in elite law reviews. *Journal of Legal Studies, 29*(1), 427–450.

Bar-Ilan, J. (2009). A closer look at the sources of informetric research. *Cybermetrics, 13*(1), p. 4.

Bartley, W. W. (1984). *The retreat to commitment* (2nd ed.). La Salle: Open Court.

Bayer, A. E., & Folger, J. (1966). Some correlates of a citation measure of productivity in science. *Sociology of Education, 39*(4), 381–390.

Borgman, C. L., & Furner, J. (2002). Scholarly communication and bibliometrics [Review]. *Annual Review of Information Science and Technology, 36*, 3–72.

Bornmann, L. (2011). Scientific peer review. *Annual Review of Information Science and Technology, 45*, 199–245.

Bornmann, L., & Daniel, H.-D. (2005a). Criteria used by a peer review committee for selection of research fellows – a boolean probit analysis. *International Journal of Selection and Assessment, 13*(4), 296–303.

Bornmann, L., & Daniel, H.-D. (2005b). Selection of research fellowship recipients by committee peer review. Analysis of reliability, fairness and predictive validity of Board of Trustees' decisions. *Scientometrics, 63*(2), 297–320.

Bornmann, L., & Daniel, H.-D. (2006). Selecting scientific excellence through committee peer review – a citation analysis of publications previously published to approval or rejection of post-doctoral research fellowship applicants. *Scientometrics, 68*(3), 427–440.

Bornmann, L., & Daniel, H.-D. (2008a). The effectiveness of the peer review process: Inter-referee agreement and predictive validity of manuscript refereeing at *Angewandte Chemie*. *Angewandte Chemie. International Edition, 47*(38), 7173–7178. doi:10.1002/anie.200800513.

Bornmann, L., & Daniel, H.-D. (2008b). Selecting manuscripts for a high impact journal through peer review: A citation analysis of communications that were accepted by *Angewandte Chemie International Edition*, or rejected but published elsewhere. *Journal of the American Society for Information Science and Technology, 59*(11), 1841–1852. doi:10.1002/asi.20901.

Bornmann, L., & Daniel, H.-D. (2008c). What do citation counts measure? A review of studies on citing behavior. *Journal of Documentation, 64*(1), 45–80. doi:10.1108/002204108108 44150.

Bornmann, L., & Daniel, H.-D. (2009a). The luck of the referee draw: The effect of exchanging reviews. *Learned Publishing, 22*(2), 117–125. doi:10.1087/2009207.

Bornmann, L., & Daniel, H.-D. (2009b). Universality of citation distributions. A validation of Radicchi et al.'s relative indicator $c_f = c/c_0$ at the micro level using data from chemistry. *Journal of the American Society for Information Science and Technology, 60*(8), 1664–1670.

Bornmann, L., & Daniel, H.-D. (2010). The manuscript reviewing process – empirical research on review requests, review sequences and decision rules in peer review. *Library & Information Science Research, 32*(1), 5–12. doi:10.1016/j.lisr.2009.07.010.

Bornmann, L., Wallon, G., & Ledin, A. (2008). Does the committee peer review select the best applicants for funding? An investigation of the selection process for two European Molecular Biology Organization programmes. *PLoS One, 3*(10), e3480.

Bornmann, L., Marx, W., Schier, H., Thor, A., & Daniel, H.-D. (2010). From black box to white box at open access journals: Predictive validity of manuscript reviewing and editorial decisions at *Atmospheric Chemistry and Physics*. *Research Evaluation, 19*(2), 81–156.

Bornstein, R. F. (1991). The predictive validity of peer-review: A neglected issue. *The Behavioral and Brain Sciences, 14*(1), 138–139.

Boyack, K. W., & Klavans, R. (2005). Predicting the importance of current papers. In P. Ingwersen & B. Larsen (Eds.), *Proceedings of the 10th International Conference of the International Society for Scientometrics and Informetrics* (Vol. 1, pp. 335–342). Stockholm: Karolinska University Press.

Buela-Casal, G., Gutiérrez-Martínez, O., Bermúdez-Sánchez, M., & Vadillo-Muñoz, O. (2007). Comparative study of international academic rankings of universities. *Scientometrics, 71*(3), 349–365.

Butler, D. (2007). Academics strike back at spurious rankings. *Nature, 447*(7144), 514–515.

Carpenter, M. P., & Narin, F. (1981). The adequacy of the Science Citation Index (SCI) as an indicator of international scientific activity. *Journal of the American Society for Information Science, 32*(6), 430–439.

Carter, G. (1982). *What we know and do not know about the peer review system* (Rand Report N-1878-RC/NIH). Santa Monica: RAND Corporation.

Castellano, C., & Radicchi, F. (2009). On the fairness of using relative indicators for comparing citation performance in different disciplines [Article]. *Archivum Immunologiae Et Therapiae Experimentalis, 57*(2), 85–90. doi:10.1007/s00005-009-0014-0.

Cole, S. (1992). *Making science: Between nature and society*. Cambridge: Harvard University Press.

Cole, J. R., & Cole, S. (1973). *Social stratification in science*. Chicago: The University of Chicago Press.

Collins, H. (2004). *Gravity's shadow: The search for gravitational waves*. Chicago: The University of Chicago Press.

Cozzens, S. E. (1985). Comparing the sciences – Citation context analysis of papers from neuropharmacology and the sociology of science. *Social Studies of Science, 15*(1), 127–153.
Cronin, B. (1984). *The citation process: The role and significance of citations in scientific communication*. Oxford: Taylor Graham.
Cronin, B. (2005), *The hand of science*. Academic Writing and its Rewards, Scarecrow Press, Lanham, MD.
Daniel, H.-D. (1993). *Guardians of science: Fairness and reliability of peer review*. Weinheim: Wiley-VCH.
Daniel, H.-D. (2005). Publications as a measure of scientific advancement and of scientists' productivity. *Learned Publishing, 18*, 143–148.
de Vries, D. R., Marschall, E. A., & Stein, R. A. (2009). Exploring the peer review process: What is it, does it work, and can it be improved? *Fisheries, 34*(6), 270–279.
Eckes, T. (2004). Rater agreement and rater severity: A many-faceted Rasch analysis of performance assessments in the "Test Deutsch als Fremdsprache" (TestDaF). *Diagnostica, 50*(2), 65–77.
Ehses, I. (2004). By scientists, for scientists. The Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – and how it functions. *B.I.F. Futura, 19*, 170–177.
Eichorn, P., & Yankauer, A. (1987). Do authors check their references – a survey of accuracy of references in 3 public-health journals. *American Journal of Public Health, 77*(8), 1011–1012.
Eisenhart, M. (2002). The paradox of peer review: Admitting too much or allowing too little? *Research in Science Education, 32*(2), 241–255.
Enserink, M. (2001). Peer review and quality: A dubious connection? *Science, 293*(5538), 2187–2188.
Enserink, M. (2007). Who ranks the university rankers? *Science, 317*(5841), 1026–1028. doi:10.1126/science.317.5841.1026.
Ernst, E., Saradeth, T., & Resch, K. L. (1993). Drawbacks of peer review. *Nature, 363*(6427), 296.
Evans, J. T., Nadjari, H. I., & Burchell, S. A. (1990). Quotational and reference accuracy in surgical journals – A continuing peer-review problem. *Journal of the American Medical Association, 263*(10), 1353–1354.
Feist, G. J. (2006). *The psychology of science and the origins of the scientific mind*. New Haven: Yale University Press.
Figueredo, E. (2006). The numerical equivalence between the impact factor of journals and the quality of the articles. *Journal of the American Society for Information Science and Technology, 57*(11), 1561.
Fiske, D. W., & Fogg, L. (1990). But the reviewers are making different criticisms of my paper – Diversity and uniqueness in reviewer comments. *The American Psychologist, 45*(5), 591–598.
Fok, D., & Franses, P. H. (2007). Modeling the diffusion of scientific publications. *Journal of Econometrics, 139*(2), 376–390. doi:10.1016/j.jeconom.2006.10.021.
Fox, T. (1965). *Crisis in communication: The functions and future of medical publication*. London: Athlone Press.
Garfield, E. (1970). Citation indexing for studying science. *Nature, 227*(5259), 669–671.
Garfield, E. (1972). Citation analysis as a tool in journal evaluation: Journals can be ranked by frequency and impact of citations for science policy studies. *Science, 178*(4060), 471–479.
Garfield, E. (1998). Random thoughts on citationology. Its theory and practice [Article]. *Scientometrics, 43*(1), 69–76.
Garfield, E. (2002). Highly cited authors. *Scientist, 16*(7), 10.
Garfield, E., & Welljamsdorof, A. (1992). Citation data – Their use as quantitative indicators for science and technology evaluation and policy-making. *Current Contents, 49*, 5–13.
Geisler, E. (2001). The mires of research evaluation. *The Scientist, 15*(10), 39.
Godlee, F., & Dickersin, K. (2003). Bias, subjectivity, chance, and conflict of interest. In F. Godlee & J. Jefferson (Eds.), *Peer review in health sciences* (2nd ed., pp. 91–117). London: BMJ Publishing Group.
Goodman, S. N., Berlin, J., Fletcher, S. W., & Fletcher, R. H. (1994). Manuscript quality before and after peer review and editing at *Annals of Internal Medicine*. *Annals of Internal Medicine, 121*(1), 11–21.

Hames, I. (2007). *Peer review and manuscript management of scientific journals: Guidelines for good practice*. Oxford: Blackwell.
Hemlin, S. (1996). Research on research evaluations. *Social Epistemology, 10*(2), 209–250.
Hojat, M., Gonnella, J. S., & Caelleigh, A. S. (2003). Impartial judgment by the "gatekeepers" of science: Fallibility and accountability in the peer review process. *Advances in Health Sciences Education, 8*(1), 75–96.
Hornbostel, S., Böhmer, S., Klingsporn, B., Neufeld, J., & von Ins, M. (2009). Funding of young scientist and scientific excellence. *Scientometrics, 79*(1), 171–190.
Hudson, J. (2007). Be known by the company you keep: Citations – quality or chance? *Scientometrics, 71*(2), 231–238.
Hurt, C. D. (1987). Conceptual citation differences in science, technology, and social sciences literature. *Information Processing and Management, 23*(1), 1–6.
Jayasinghe, U. W., Marsh, H. W., & Bond, N. (2001). Peer review in the funding of research in higher education: The Australian experience. *Educational Evaluation and Policy Analysis, 23*(4), 343–346.
Jensen, P., Rouquier, J. B., & Croissant, Y. (2009). Testing bibliometric indicators by their prediction of scientists promotions. [Article]. *Scientometrics, 78*(3), 467–479. doi:10.1007/s11192-007-2014-3.
Judge, T., Cable, D., Colbert, A., & Rynes, S. (2007). What causes a management article to be cited – article, author, or journal? *The Academy of Management Journal (AMJ), 50*(3), 491–506.
Kellogg, D. (2006). Toward a post-academic science policy: Scientific communication and the collapse of the Mertonian norms. *International Journal of Communications Law & Policy, 11*, IJCLP Web-Doc 1-11-2006.
Kellsey, C., & Knievel, J. E. (2004). Global English in the humanities? A longitudinal citation study of foreign-language use by humanities scholars. *College & Research Libraries, 65*(3), 194–204.
King, J. (1987). A review of bibliometric and other science indicators and their role in research evaluation. *Journal of Information Science, 13*(5), 261–276.
Knorr-Cetina, K. (1981). *The manufacture of knowledge: An essay on the constructivist and contextual nature of science*. Oxford: Pergamon.
Knorr-Cetina, K. (1991). Merton sociology of science: The first and the last sociology of science. *Contemporary Sociology, 20*(4), 522–526.
Kuhn, T. S. (1962). *The structure of scientific revolutions*. Chicago: University of Chicago Press.
LaFollette, M. C. (1992). *Stealing into print: Fraud, plagiarism and misconduct in scientific publishing*. Berkeley: University of California Press.
Laloë, F., & Mosseri, R. (2009). Bibliometric evaluation of individual researchers: Not even right… not even wrong! *Europhysics News, 40*(5), 26–29.
Langfeldt, L. (2006). The policy challenges of peer review: Managing bias, conflict of interests and interdisciplinary assessments. *Research Evaluation, 15*(1), 31–41.
Lansingh, V. C., & Carter, M. J. (2009). Does Open Access in ophthalmology affect how articles are subsequently cited in research? [Article]. *Ophthalmology, 116*(8), 1425–1431. doi:10.1016/j.ophtha.2008.12.052.
Larivière, V., & Gingras, Y. (2010). The impact factor's Matthew Effect: A natural experiment in bibliometrics. *Journal of the American Society for Information Science and Technology, 61*(2), 424–427.
Latour, B., & Woolgar, S. (1979). *Laboratory life: The social construction of scientific facts*. London: Sage.
Lawani, S. M. (1977). The professional literature used by American and French agronomists and the implications for agronomic education. *Journal of Agronomic Education, 6*, 41–46.
Lawani, S. M. (1986). Some bibliometric correlates of quality in scientific research. *Scientometrics, 9*(1–2), 13–25.
Lawrence, S. (2001). Free online availability substantially increases a paper's impact. *Nature, 411*(6837), 521–521.

Leimu, R., & Koricheva, J. (2005). What determines the citation frequency of ecological papers? *Trends in Ecology & Evolution, 20*(1), 28–32.
Lewison, G. (1998). Gastroenterology research in the United Kingdom: Funding sources and impact. *Gut, 43*(2), 288–293.
Lienert, G. A. (1987). *Schulnoten-evaluation*. Frankfurt am Main: Athenäum.
Lindsey, D. (1989). Using citation counts as a measure of quality in science. Measuring what's measurable rather than what's valid. *Scientometrics, 15*(3–4), 189–203.
Liu, Z. M. (1997). Citation theories in the framework of international flow of information: New evidence with translation analysis. *Journal of the American Society for Information Science, 48*(1), 80–87.
Long, J. S., & Fox, M. F. (1995). Scientific careers – Universalism and particularism. *Annual Review of Sociology, 21*, 45–71.
Lundberg, J. (2007). Lifting the crown – citation *z*-score. *Journal of Informetrics, 1*(2), 145–154.
Mählck, P., & Persson, O. (2000). Socio-bibliometric mapping of intra-departmental networks. *Scientometrics, 49*(1), 81–91.
Mallard, G., Lamont, M., & Guetzkow, J. (2009). Fairness as appropriateness: Negotiating epistemological differences in peer review. *Science Technology Human Values, 34*(5), 573–606. doi:10.1177/0162243908329381.
Martin, B. R., & Irvine, J. (1983). Assessing basic research – Some partial indicators of scientific progress in radio astronomy. *Research Policy, 12*(2), 61–90.
McClellan, J. E. (2003). *Specialist control – The publications committee of the Academie Royal des Sciences (Paris) 1700–1793* (Transactions of the American Philosophical Society, Vol. 93). Philadelphia: American Philosophical Society.
McDonald, J. D. (2007). Understanding journal usage: A statistical analysis of citation and use. *Journal of the American Society for Information Science and Technology, 58*(1), 39–50.
McDonald, R. J., Cloft, H. J., & Kallmes, D. F. (2009). Fate of manuscripts previously rejected by the *American Journal of Neuroradiology*: A follow-up analysis. *American Journal of Neuroradiology, 30*(2), 253–256. doi:10.3174/Ajnr.A1366.
Melin, G., & Danell, R. (2006). The top eight percent: Development of approved and rejected applicants for a prestigious grant in Sweden. *Science and Public Policy, 33*(10), 702–712.
Menou, M. J. (1983). Cultural barriers to the international transfer of information. *Information Processing and Management, 19*(3), 121–129.
Merisotis, J., & Sadlak, J. (2005). Higher education rankings: Evolution, acceptance, and dialogue. *Higher Education in Europe, 30*(2), 97–101.
Merton, R. K. (1942). Science and technology in a democratic order. *Journal of Legal and Political Sociology, 1*, 115–126.
Merton, R. K. (1957). Priorities in scientific discovery: A chapter in the sociology of science. *American Sociological Review, 22*(6), 635–659. doi:10.2307/2089193.
Merton, R. K. (1968). The Matthew effect in science. *Science, 159*(3810), 56–63.
Merton, R. K. (1973). *The sociology of science: Theoretical and empirical investigations*. Chicago: University of Chicago Press.
Merton, R. K. (1988). The Matthew effect in science, II: Cumulative advantage and the symbolism of intellectual property. *ISIS, 79*(4), 606–623.
Moed, H. (2008). UK Research Assessment Exercises: Informed judgments on research quality or quantity? *Scientometrics, 74*(1), 153–161.
Narin, F. (1976). *Evaluative bibliometrics: The use of publication and citation analysis in the evaluation of scientific activity*. Cherry Hill: Computer Horizons.
National Academy of Sciences. (2006). *Beyond bias and barriers: Fulfilling the potential of women in academic science and engineering*. Washington: The National Academies Press.
Nicolaisen, J. (2002). The J-shaped distribution of citedness. *Journal of Documentation, 58*(4), 383–395.
Nicolaisen, J. (2003). The social act of citing: Towards new horizons in citation theory. In The social act of citing: towards new horizons in citation theory. *Proceedings of the 66th ASIST Annual Meeting*, 12–20.

Nicolaisen, J. (2007). Citation analysis. *Annual Review of Information Science and Technology, 41*, 609–641.

Oppenheim, C. (1995). The correlation between citation counts and the 1992 research assessment exercise ratings for British library and information science university departments. *Journal of Documentation, 51*(1), 18–27.

Oppenheim, C. (1997). The correlation between citation counts and the 1992 research assessment exercise ratings for British research in genetics, anatomy and archaeology. *Journal of Documentation, 53*(5), 477–487.

Opthof, T., Furstner, F., van Geer, M., & Coronel, R. (2000). Regrets or no regrets? No regrets! The fate of rejected manuscripts. *Cardiovascular Research, 45*(1), 255–258.

Owen, R. (1982). Reader bias. *Journal of the American Medical Association, 247*(18), 2533–2534.

Peters, H. P. F., & van Raan, A. F. J. (1994). On determinants of citation scores – A case study in chemical engineering. *Journal of the American Society for Information Science, 45*(1), 39–49.

Pierie, J. P. E. N., Walvoort, H. C., & Overbeke, A. J. P. M. (1996). Readers' evaluation of effect of peer review and editing on quality of articles in the *Nederlands Tijdschrift voor Geneeskunde*. *Lancet, 348*(9040), 1480–1483.

Popper, K. R. (1961). *The logic of scientific discovery* (2nd ed.). New York: Basic Books.

Pruthi, S., Jain, A., Wahid, A., Mehra, K., & Nabi, S. A. (1997). Scientific community and peer review system – A case study of a central government funding scheme in India. *Journal of Scientific and Industrial Research, 56*(7), 398–407.

Radicchi, F., Fortunato, S., & Castellano, C. (2008). Universality of citation distributions: Toward an objective measure of scientific impact. *Proceedings of the National Academy of Sciences of the United States of America, 105*(45), 17268–17272. doi:10.1073/pnas.0806977105.

Reinhart, M. (2009). Peer review of grant applications in biology and medicine. Reliability, fairness, and validity. *Scientometrics, 81*(3), 789–809. doi:10.1007/s11192-008-2220-7.

Research Evaluation and Policy Project. (2005). *Quantitative indicators for research assessment – A literature review* (REPP discussion paper 05/1). Canberra: Research Evaluation and Policy Project, Research School of Social Sciences, The Australian National University.

Ross, P. F. (1980). *The sciences' self-management: Manuscript refereeing, peer review, and goals in science*. Lincoln: The Ross Company.

Sandström, U., Wadskog, D., & Karlsson, S. (2005). Research institutes and universities: Does collaboration pay? In P. Ingwersen & B. Larsen (Eds.), *Proceedings of the 10th International Conference of the International Society for Scientometrics and Informetrics* (Vol. 2, pp. 690–691). Stockholm, Sweden: Karolinska University Press.

Schmelkin, L. (2006). *Weaknesses of peer reviewing and peer refereeing*. Paper presented at the first international conference on Knowledge Communication and Peer Reviewing, Orlando.

Schneider, J. W. (2009). An outline of the bibliometric indicator used for performance-based funding of research institutions in Norway. *European Political Science, 8*(3), 364–378. doi:10.1057/Eps.2009.19.

Shadbolt, N., Brody, T., Carr, L., & Harnad, S. (2006). The Open Research Web: A preview of the optimal and the inevitable. In N. Jacobs (Ed.), *Open access: Key strategic, technical and economic aspects* (pp. 195–208). Oxford: Chandos.

Sharp, D. W. (1990). What can and should be done to reduce publication bias? The perspective of an editor. *Journal of the American Medical Association, 263*(10), 1390–1391.

Shashok, K. (2005). Standardization vs diversity: How can we push peer review research forward? *Medscape General Medicine, 7*(1), 11.

Shatz, D. (2004). *Peer review: A critical inquiry*. Lanham: Rowman & Littlefield.

Sismondo, S. (1993). Some social constructions [Article]. *Social Studies of Science, 23*(3), 515–553.

Smith, L. C. (1981). Citation analysis. *Library Trends, 30*(1), 83–106.

Smith, R. (2006). Peer review: A flawed process at the heart of science and journals. [Article]. *Journal of the Royal Society of Medicine, 99*(4), 178–182.

Stamps, A. E. (1997). Advances in peer review research: An introduction. *Science and Engineering Ethics, 3*(1), 3–10.

Steele, C., Butler, L., & Kingsley, D. (2006). The Publishing imperative: The pervasive influence of publication metrics. *Learned Publishing, 19*(4), 277–290.

Stewart, J. A. (1983). Achievement and ascriptive processes in the recognition of scientific articles. *Social Forces, 62*(1), 166–189.

Tijssen, R. J. W., van Leeuwen, T. N., & van Raan, A. F. J. (2002). *Mapping the scientific performance of German medical research. An international comparative bibliometric study*. Stuttgart: Schattauer.

UK Office of Science and Technology. (2006). *Science and innovation investment framework 2004–2014: Next steps*. London: UK Office of Science and Technology.

Unver, B., Senduran, M., Kocak, F. U., Gunal, I., & Karatosun, V. (2009). Reference accuracy in four rehabilitation journals [Article]. *Clinical Rehabilitation, 23*(8), 741–745. doi:10.1177/0269215508102968.

van den Besselaar, P., & Leydesdorff, L. (2007). *Past performance as predictor of successful grant applications: A case study*. Den Haag: Rathenau Instituut.

van Raan, A. F. J. (1996). Advanced bibliometric methods as quantitative core of peer review based evaluation and foresight exercises. *Scientometrics, 36*(3), 397–420.

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

van Raan, A. F. J. (2005b). For your citations only? Hot topics in bibliometric analysis. *Measurement: Interdisciplinary Research and Perspectives, 3*(1), 50–62.

Weingart, P. (2005a). Das Ritual der Evaluierung und die Verführbarkeit. In P. Weingart (Ed.), *Die Wissenschaft der Öffentlichkeit: Essays zum Verhältnis von Wissenschaft, Medien und Öffentlichkeit* (pp. 102–122). Weilerswist: Velbrück.

Weingart, P. (2005b). Impact of bibliometrics upon the science system: Inadvertent consequences? *Scientometrics, 62*(1), 117–131.

Weller, A. C. (2002). *Editorial peer review: Its strengths and weaknesses*. Medford: Information Today.

White, H. D. (2001). Authors as citers over time. *Journal of the American Society for Information Science and Technology, 52*(2), 87–108.

Whitley, R., & Gläser, J. (Eds.). (2007). *The changing governance of the sciences: The advent of research evaluation systems*. Dordrecht: Springer.

Wiley, S. (2008). Peer review isn't perfect … But it's not a conspiracy designed to maintain the status quo. *The Scientist, 22*(11), 31.

Wilson, J. D. (1978). Peer review and publication. *The Journal of Clinical Investigation, 61*(4), 1697–1701.

Wood, F. Q., & Wessely, S. (2003). Peer review of grant applications: A systematic review. In F. Godlee & T. Jefferson (Eds.), *Peer review in health sciences* (2nd ed., pp. 14–44). London: BMJ Books.

Ziman, J. (2000). *Real science: What it is, and what it means*. Cambridge: Cambridge University Press.

Chapter 9
Measuring Teaching Performance

Keith Trigwell

9.1 Global University Ranking Systems

This chapter is concerned with the place of teaching and learning in university rankings – both global and national. As a key component of core university business, teaching and learning should be a key component of any ranking exercise, but this is mostly not so. For example, of the two principal global university rankings, only the QS Times Higher Education 'World University Rankings' includes an attempt (and a mediocre one at that) to measure aspects of teaching and learning. The 'Academic Ranking of World Universities' compiled by the Shanghai Jiao Tong University Institute of Higher Education (SJTI) group focuses mainly on research performance. At the national level, the inclusion of indicators of teaching and learning in ranking calculations are more common, but the focus is still heavily on research indicators.

Such a focus is understandable, for two reasons, as Liu and Cheng (2005) suggest. First, international research indicators are well-established and are commonly accepted. Second, agreeing on, and implementing indicators of teaching performance tends to be difficult 'owing to the huge differences between universities and the large variety of countries, and because of the technical difficulties inherent in obtaining internationally comparable data' (Liu and Cheng 2005: 133). Similar conclusions are reached by other commentators. A repeated difficulty is that no ranking or quality-assessment system has been able to generate data based on measures of the 'value added' during the educational process; so, few focus on teaching and learning at all (Dill and Soo 2005: 503–505). As Altbach (2006) states, 'There are, in fact, no widely accepted methods for measuring teaching quality, and assessing the impact of education on students is so far an unexplored area as well' (Marginson and van der Wende 2007: 2; Guarino et al. 2005: 149).

In the QS Times Higher Education ranking, a high value is placed on institutional reputation and on the level of 'internationalisation' of Higher Education

K. Trigwell (✉)
Institute for Teaching and Learning, University of Sydney, Sydney, Australia
e-mail: Keith.trigwell@sydney.edu.au

J.C. Shin et al. (eds.), *University Rankings*, The Changing Academy – The Changing Academic Profession in International Comparative Perspective 3,
DOI 10.1007/978-94-007-1116-7_9, © Springer Science+Business Media B.V. 2011

Institutions (HEIs), and in the outcome, the rankings tend to favour HEIs with a strong presence in the degree market. A total of 40% of the Times index is composed of an opinion survey of academics around the world, and another 10% of a survey of global employers. There are two internationalisation indicators: the proportion of students who are international (5%) and the proportion of staff who are international (5%). Another 20% is determined by the student–staff ratio which is treated as a proxy for teaching 'quality'. The remaining 20% of the Times index is composed of research citations per staff member using the Thomson and ISI database (The Times Higher Education 2010).

The ratio of students to teaching staff (student–staff ratio) is probably the most accessible and widely used measure of teaching in the world. Research into student learning shows that there is a significant negative correlation between class size and quality of learning, and as noted in the QS Times HES literature, it does at least provide an indication that the institution in question has sufficient staff to teach its students (Sowter 2009). But there would be few people, if any, who would consider that the quality of university teaching (and learning) can be assessed using this measure alone. Of the remaining QS Times HES indicators, few have any clear connection to teaching quality. Reputation, as assessed by academic peers, is likely to be heavily weighted towards research, and international student demand is also mostly about research reputation.

Given that teaching and learning is not a key element of the global rankings of either international system, it is difficult to understand the claim made by both that potential students and parents are their key users (Stromquist 2007).

9.2 National Ranking Systems

In a recent analysis of the impact of ranking on institutional decision making, the Institute for Higher Education Policy in the USA noted that:

> At least 40 different nations now have some form of rankings that are regularly published, with more added each year. The criteria vary significantly; the best known and most influential ranking schemes evaluate institutions primarily on the basis of academic factors, but some ranking systems use the number of new patents acquired, climate for female students, environmentally friendly practices, and a host of other characteristics. (Institute for Higher Education Policy 2009: 5)

In the USA, the journal *US News and World Reports* started the annual publication of 'America's Best Colleges' in 1983, and other countries also quickly established their own national rankings using their own measures. For example, national rankings are now found in universities in Germany (Federkeil 2002); Australia and New Zealand (Clarke 2005); China (Liu and Liu 2005); Japan (Yonezawa et al. 2002); Spain (De Miguel et al. 2005); Latin-America (Ledesma 2002); Britain (Eccles 2002); Poland (Perspektywy 2002); and Russia (Filinov and Ruchkina 2002).

Among the academic factors other than research that feature in these systems are, rating of teachers by students, students' study times, accessibility of teachers,

availability of class discussion, library, study abroad, best classroom experience (The Princeton Review 2010), and students' perceptions of teaching quality, attainment of generic skills and overall satisfaction with the quality of their programmes, as well as academic progress, retention and employment/study following graduation (Department of Education, Science and Training, Australian Government 2004). The seven categories and the corresponding percentage weightings used by the US News and World Report are: peer assessment (25%); retention (20% in national universities and liberal arts colleges and 25% in master's and baccalaureate colleges); faculty resources (20%); student selectivity (15%); financial resources (10%); graduation rate performance (5%; only in national universities and liberal arts colleges) and alumni giving rate (5%) (Morse and Flemingan 2009).

While these examples illustrate the range of indicators that are being used nationally, the question remains as to whether they are indicators of what is now recognised as good teaching (or learning) and whether they capture variation in the quality of education provided or achieved. This question is addressed later in the chapter. The focus of the next few sections is on the research that has led to the practices described above.

9.3 Is Research a Proxy for Teaching Quality?

It has been argued by some (Benowski 1991; Taylor 2003; Yair 2008) that because there is a strong link between university research and teaching, that universities that are strong in research will be strong in teaching, and those indicators of research strength are also indicators of strength in teaching. From this perspective, the teaching–research relationship is bidirectional, and one can influence and enhance the other. It is underpinned by the belief that by researching the subject matters they teach, academic staff can gain and maintain passion, which is a prerequisite for excellence in teaching. In addition, the practice of teaching can foster academics to gain a deeper understanding of the field, and absorb novel ideas from their students, rendering them better researchers (Taylor 2003).

Contrary to the view above, other scholars contend that teaching excellence and research excellence are 'in direct conflict, compete for academic attention and only one of them is rewarded' (Gibbs 1995: 149). Gibbs (1995) quotes some empirical studies on teaching and research relations to corroborate his claims. For instance, a survey of Higher Education Institutions in the UK conducted by the Oxford Centre for Staff Development found that only 12% of promotion decisions were made solely on the basis of teaching excellence; and 38% of institutions made promotion decisions on grounds that did not include quality of teaching. The conflicting relation between teaching and research is also reported in other countries (e.g. Astin and Chang 1995) and in specific fields of study (e.g. Patrick and Stanley 1998). Astin and Chang (1995), using a survey method, discovered that no institutions in the USA (in their sample) have a balanced orientation towards both students (learning) and research.

Patrick and Stanley (1998) failed to show any connections between high quality of teaching and excellence of research in business fields of study. Based on a meta-analysis of 58 articles, Hattie and Marsh (1996) had earlier reported a near zero (0.06) correlation between indicators of teaching and research, using various measures of teaching quality (e.g. student evaluation, peer ratings) and of research output (e.g. citations, productivity, quality). And in a later paper, Marsh and Hattie (2002) provide evidence confirming the earlier results, noting a correlation of only 0.02 between global teaching and global research factors. These and other studies have led Dill and Soo (2005: 507) to conclude that 'empirical research … suggests that the correlation between research productivity and undergraduate instruction is very small and teaching and research appear to be more or less independent activities'.

This brief review suggests that there is certainly insufficient evidence available to support the idea that research quality can be used as a proxy for teaching quality. So, what approaches are available to measure teaching quality?

9.4 Measurements of Quality Teaching

Although assessing quality teaching remains a contentious issue among educational researchers, it is agreed upon that the choice of indicators to measure teaching quality is crucial, and it is essential to select valid indicators rather than just practical ones.

Brooks (2005) has proposed the following choice of indicators for teaching:

1. Using student questionnaires
   Employing student evaluation of teaching seems logical since teaching quality is essentially student-centred and aims at improving student learning. However, many teachers perceive student evaluation is more or less biased, thus they give little credence to the responses obtained from students.
2. Peer in-class evaluations
   Peer in-class evaluations has an obvious advantage in that it focuses more on teaching process rather than outcome, yet it may hinder teaching innovations under the influence of conservative judgement held by peer evaluators. It is also only ever an observation of a small part of teaching.
3. Using teaching portfolios
   Using teaching portfolios to indicate quality teaching seems fair, yet it is not easy to operationalise as each field has their own practice, making comparison across fields effectively impossible.

Since teaching is a means to the student learning end, an alternate approach is to focus on the students' programmes, their experience and outcome. Measurements of these areas are seen by Brooks to be comprised of four main categories, namely, programme characteristics, programme effectiveness, student satisfaction and student outcomes.

1. Programme characteristics include counts of the degrees awarded, amount of student financial support available and quality of students entering the university.
2. Programme effectiveness is a measure of the timelines of completion of students.
3. Student satisfaction is often measured by national survey or by examining students' engagement and career preparation.
4. Student outcomes are often operationalised by either determining students' career path and job satisfaction or their generic skills, such as critical thing, analytical reasoning and written communication (Brooks 2005).

The combination of the teaching indicators used in national rankings with these seven areas constitutes a comprehensive array of what might be the basis of a teaching index. But how close do these elements or even the full array, match what is now known about high quality university teaching?

9.5 What Is High Quality Teaching?

A common explanation as to why there are so few indicators of good teaching in global ranking systems is that high quality teaching (and learning) is too difficult to define and to measure. A part of this explanation is no longer valid. It is now quite clear what good teaching is, and what constitutes poor teaching. International research studies consistently point to a set of factors associated with good teaching (and learning) and these are addressed in more detail below. However, there are still scholars who hold the belief that *quality* teaching can never be fully understood since teaching itself is a never-ending process (Argyris and Schön 1974; Hau 1996). For instance, Hau (1996) contends that in order to ensure quality teaching in higher education, a never-ending process should be undergone to eliminate defects in quality.

As Biggs (2001) suggested, 'quality' is unarguably a multi-layered concept, which can be understood as an outcome, a property, or a process. Quality teaching also encompasses multiple meaning. The contentiousness of 'quality' also stems from its 'stakeholder-relative' characteristics, as suggested by Tam (2001), who proposes that all stakeholders hold their own perspective on what is meant by quality in education. This idea is also supported by Harvey and his associates, who agree that different ways of defining quality in higher education results from different stakeholders in higher education (Harvey et al. 1992). According to Harvey et al. (1992), 'stakeholders' include students, teaching and non-teaching staff, government and funding agencies, assessors and even the community.

Harvey and Green (1993) describe four kinds of definitions of quality teaching according to its history of development. Traditionally, quality teaching is a conception closely related to 'excellence', which is still a dominant concept among many tertiary educational institutions. More recently, quality teaching is associated with 'value for money'. Thirdly, quality teaching can also be defined as the efficiency and effectiveness of fitness for purpose, which suggests that a quality educational

provider needs to engage students to acquire knowledge effectively. A final conceptualisation of quality has to do with 'transforming' students' perceptions of applying their knowledge to solve real world problems.

Quality teaching may be hard to define, and excellent teaching may be difficult to achieve, but researchers do generally agree that the following list includes those qualities that are essential to good teaching (Ramsden et al. 1995):

- Good teachers are also good learners; for example, they learn through their own reading, by participating in a variety of professional development activities, by listening to their students, by sharing ideas with their colleagues and by reflecting on classroom interactions and students' achievements. Good teaching is therefore dynamic, reflective and constantly evolving.
- Good teachers display enthusiasm for their subject, and a desire to share it with their students.
- Good teachers recognise the importance of context, and adapt their teaching accordingly; they know how to modify their teaching strategies according to the particular students, subject matter and learning environment.
- Good teachers encourage learning for understanding and are concerned with developing their students' critical thinking skills, problem-solving skills and problem-approach behaviours.
- Good teachers demonstrate an ability to transform and extend knowledge, rather than merely transmitting it; they draw on their knowledge of their subject, their knowledge of their learners and their general pedagogical knowledge to transform the concepts of the discipline into terms that are understandable to their students. In other words, they display what Shulman has termed 'pedagogical content knowledge' (Shulman 1987, 1988).
- Good teachers set clear goals, use valid and appropriate assessment methods and provide high quality feedback to their students.
- Good teachers show respect for their students; they are interested in both their professional and their personal growth, encourage their independence and sustain high expectations of them. (p. 24)

9.5.1 Scholarship of Teaching

It has been argued by Boyer (1990) that the elements of university teaching should be rendered a higher status and that one way to do that is to encompass the 'scholarship of teaching' as one of the four types of scholarships in the academy (teaching, discovery, integration and application). Trigwell et al. (2000) have reviewed the meaning of scholarly teaching and presented it as an achievement in four broad dimensions (such as communication and ways of thinking about teaching). A generic approach to the assessment of these scholarships, including teaching, is described by Glassick et al. (1997) in work, building on that of Boyer. Using the concept of scholarship, Glassick and colleagues proposed that academics practising

- Clear Goals - Does the scholar state the basic purposes of his or her work clearly? Does the scholar define objectives that are realistic and achievable? Does the scholar identify important questions in the field?
- Adequate Preparation - Does the scholar show an understanding of existing scholarship in the field? Does the scholar bring the necessary skills to her or his work? Does the scholar bring together the resources necessary to move the project forward?
- Appropriate Methods - Does the scholar use methods appropriate to the goals? Does the scholar apply effectively the methods selected? Does the scholar modify procedures in response to changing circumstances?
- Significant Results - Does the scholar achieve the goals? Does the scholar's work add consequentially to the field? Does the scholar's work open additional areas for further exploration?
- Effective Presentation - Does the scholar use a suitable style and effective organisation to present his or her work? Does the scholar use appropriate forums for communicating work to its intended audiences? Does the scholar present her or his message with clarity and integrity?
- Reflective Critique - Does the scholar critically evaluate his or her own work? Does the scholar bring an appropriate breadth of evidence to her or his critique? Does the scholar use evaluation to improve the quality of future work?

(Glassick *et al*., 1997: 36).

Fig. 9.1 The Carnegie Foundation's six standards of scholarship

the 'scholarship of teaching' experience a symmetry between teaching and research through the common elements of clear goals, adequate preparation, appropriate methods, significant results, effective publication and reflective critique (Glassick et al. 1997). If these criteria are used to assess the quality of the scholarship of discovery (research), can they also be used to assess the quality of the scholarship of teaching? And if they can, is there a relation between scholarship of teaching and good teaching or high quality teaching? The symmetry between the two forms of scholarship is illustrated in Fig. 9.1.

9.5.2 Qualitative Variation in University Teaching

It is apparent from the above analysis that teaching involves much more than what happens in a classroom or on-line; it is oriented towards, and is related to, high quality student learning, and includes planning, compatibility with the context, content knowledge, being a learner and above all, a certain way of thinking about teaching and learning. It is also no longer a solo pursuit. Not only do the planning elements apply to individual subjects, they must also be seen to be part of a whole programme. And there is an alignment between teachers' teaching and student learning. Good teaching includes all these elements.

Higher education research studies suggest that there is a way of conceiving university teaching which is more strongly associated with higher quality student learning than other ways of thinking (Prosser and Trigwell 1999). Some teachers keep more of a focus on their students in their planning and their activities. These teachers tend to be teaching students who describe using a higher quality approach to their learning. Teachers adopting this approach see their role as helping their students develop and change their conceptions or world views. As a result of this thinking, their focus is on the bigger picture – an overview of the topic or how the components of the information are related to each other, and on students' prior knowledge – what students bring to the situation. Their planning and teaching methods are in alignment with this conception.

This thinking is in contrast to that of teachers who work with a view where the focus is on what they do as teachers, or on the detail – individual concepts in the syllabus or textbook, or the teachers' own knowledge structure – without acknowledgment of what students may bring to the situation or experience in the situation. They see their role as transmitting information based upon that knowledge to their students.

With respect to the concept of alignment, a teacher who holds the former conception is more likely to adopt an approach which has the student as the focus of activities. It matters more to this teacher what the student is doing and learning and experiencing than what the teacher is doing or covering. This teacher is one who encourages self-directed learning, who makes time (in 'formal teaching' time) for students to interact and to discuss the problems they encounter, who assesses to reveal transformed knowledge (not only to judge and rank students), who provokes debate (and raises and addresses the taken-for-granted issues) and who uses a lot of 'lecture' time to question students' ideas, and to develop a 'conversation' with the students.

These strategies may differ from those used by a teacher with a teacher-focused approach, but this is not always so. For example, two teachers can use the same strategy (say, buzz groups during a lecture – a buzz group is a short discussion between a small group of students, initiated by the teacher). It is the teachers' intention (aligned with their conception) that constitutes the main difference in this case. Using a student-focused approach, a teacher may see the buzz groups as a means by which students can compare their understandings of the lecture topic, and give feedback to the teacher on that understanding. In a teacher-focused approach, the teacher may see buzz groups as a way of giving her or himself a break from talking, and students a break from note-taking in a one-hour lecture. The differences in student learning, from the use of the same strategy, may be substantial.

Much of the good teaching literature, and some of the more common forms of evaluating teaching, are focused on teaching strategies, such as the clarity of explanations, or the availability of teachers for consultations. The conception-based research described above suggests that unless the teacher is using a student-focused, conceptual development conception, the emphasis on strategies may be misplaced. For example, on advice about using an online teaching strategy for a component of a course, from a student-focused, conceptual development conception

a teacher might ask the following two questions: (a) Is this strategy likely to achieve the student learning aims? and (b) What type of learning is likely to be encouraged using this strategy? From a teacher-focused perspective, the questions asked are more likely to include: (a) Is this strategy likely to be the most efficient method of dissemination? and (b) What amount of coverage is likely to be achieved using this approach?

This variation in thinking or conceptualising is a key element in variation in teaching quality (and student learning). Teaching indicators need to focus beyond how well the teacher is conducting teaching activities, and how well those activities are received by students. There is a need to also consider the nature of those activities and how they align with variation in student learning. Activities that are student-focused are more likely to align with higher quality outcomes of learning.

This analysis of teaching and learning gives some insight into the gap that exists between the ways teaching/learning is assessed, the ways it is proposed to be assessed and what researchers agree as the characteristics of good teaching/learning.

9.6 Surveys of Student Experience

Some of these elements of good teaching have been incorporated into surveys of the student learning experience. Students' descriptions of their learning experience have also been used as a proxy for learning quality, for example, in the National Student Survey (NSS) used in the UK and the Course Experience Questionnaire (CEQ) in Australia. The research underpinning the Australian Course Experience Questionnaire (Ramsden 1991) shows that there is a significant positive correlation between the quality of the outcomes of student learning and their experience of the teaching they receive, of the appropriateness of their workload, of the nature of their assessment and of the clarity of what is expected of them (Lizzio et al. 2002).

The CEQ, as currently used as a national indicator of learning experience, contains 13 items – six that make up a scale related to students' experience of the quality of teaching in their course, six on their acquisition of a range of generic graduate skills and one item on their overall satisfaction with the quality of the course they have just completed. In a symposium designed to explore the characteristics of courses that had high scores on the Australian CEQ Good Teaching Scale, most of the representatives of the Business/Commerce/Law courses, selected to present their information, were not sure why they received the positive response from students that they did. There was enormous variation in the contexts they described. The contexts included distance learning, large undergraduate courses, small postgraduate courses, and in the disciplines of accounting, marketing, business administration, management, economics and law. At first sight, there appeared to be little that the programmes as described had in common, and that could account for the common high quality experience of their students. However, what was

apparent was that the presenters themselves, who in most cases were the course coordinators, had a lot in common. They were teachers who showed that they were concerned about students, no matter what their context.

A common theme in all courses described was closely related to the idea of using the objectives of the students as the design/presentation feature. While the focus of this theme is still on the students, it is on those students who were enrolled and doing the course rather than on those who might enrol in the course. This meant that there was almost always close contact of some sort between student and teacher as perceived by the student, and it is likely that this is perceived as personal attention, even in the larger classes.

Given the large variation in these courses in so many other aspects, this perception of personal attention may be an important factor in explaining the high CEQ rating in the courses featured at the symposium, and in providing support for the use of CEQ scores in rankings. The way the comments were made in the presentations included:

- An investigation of the nature of potential students and their objectives and the use of this information in developing the curriculum.
- Dedicated staff who make time to interact with the students.
- Personal attention given to students during industrial placement.
- An intimate rural context with significant contact between staff and students.
- Small (postgraduate) classes where students felt they had personal contact.
- Use of full-time teaching staff rather than sessional, casual or part-time staff which meant staff supporting the programme were usually more contactable.
- Programmed individual consultation time with students which was based on a formula that increased available time with increasing student numbers.
- Focus on the students, rather than on the teacher and what the teacher does.

In all the cases described above, the students received more attention than they would in a 'normal' programme. While it was not all the same sort of attention (not all classes were small or had dedicated staff), it may have been what students wanted, and it may have matched the needs of the students in the course. It is likely that what came across to students was care and support in their context. The variations in context are illustrated by extremes of a student-based approach to care and support in developing student learning independence in one university versus offering care and support in negotiating student demands for a teacher-based approach to teaching in another university.

Conclusions which can be drawn from these observations are:

1. Good teaching in any one university may be achieved when students feel they have received personal attention from the staff managing their learning programme. This personal attention may come in a variety of forms depending on the context, and it may be more important to students than issues such as class size, coherent curricula and so on.
2. High CEQ scores result from a match between course and context. Using CEQ results (or any ranking system) in advertising to attract students to the course may lead to the recruitment of students who are not those who would normally

enrol and may therefore not fit well in the existing course. Given the diversity of university missions, this one observation may be sufficient grounds to argue that global rankings based on teaching are inappropriate.

In 2005, the NSS was introduced to English universities on a voluntary basis, to be completed by undergraduate students in their final year. Since then most universities have administered the survey annually. The stated aim of the survey 'is to gather feedback on the quality of students' courses, to help inform the choices of future applicants to higher education, and to contribute to public accountability' (Higher Education Academy 2010a). The survey consists of 22 items in six broad educational areas, including teaching, assessment and feedback, academic support, organisation and management, learning resources and personal development. There is also an 'overall satisfaction' question about how satisfied the respondent is with the quality of their course (Surridge 2008). In the USA, The National Survey of Student Engagement (NSSE), first launched in 1991 and distributed annually to four-year college and university undergraduate students, asks them about 'participation in programs and activities that institutions provide for their learning and personal development' (Kuh 2001; National Survey of Student Engagement 2010a). The NSSE is comprised of 42 questions measuring students' behaviours and institutional features, which can reflect students learning and personal development. The benchmarks used in the NSSE include level of academic challenge, active and collaborative learning, student–faculty interaction, enriching educational experiences and supportive campus environment (National Survey of Student Engagement 2010b). It is currently being used in voluntary (and private) ranking exercises in the USA and Canada, and interest in an Australian version is growing (Australian Council for Education Research 2010).

Brooks (2005) proposes student surveys as an important element in choice of teaching quality indicators. Of course, there are as yet no surveys of this type that are global, and even if there were, they may prove to have the same two limitations that Marsh and Cheng (2008) describe in their report on the English National Student Survey. First, that there is not substantial variation between different universities in terms of overall satisfaction; and differences between universities explain only about 2.5% of the variance (variance component based on multilevel analyses, controlling for student characteristics and discipline). However, because the number of students from each institution is so large, the differences between institutions are quite reliable (for all but a handful of universities with small numbers of respondents). Second, that there is substantially more variance explained in overall satisfaction by differences between discipline groupings within universities than by differences between universities.

Results from surveys such as the Australian Course Experience Questionnaire, the English National Student Survey and the National Survey of Student Engagement in the USA all provide opportunities for the collection of nationwide teaching/learning related data that could potentially be used in national ranking systems, but the difficulties in developing even national teaching indicators are illustrated in the following Australian case study (Prosser M., 2009, Personal communication).

For over 5 years in Australia, the Course Experience Questionnaire has been used as the principal component of a teaching/learning performance-based funding system. In April 2004, the Australian Government released its discussion paper on its Learning and Teaching Performance Fund (Department of Education, Science and Training, Australian Government 2004). It canvassed various ways of assessing excellence based upon performance indicators, peer review and value added approaches. In the end, it adopted a two-stage process.

In stage 1, institutions had to meet a number of more qualitative criteria to be eligible for the second stage. These criteria included such things as evidence of systematic support for improving teaching and learning. Having met these conditions, institutions were deemed eligible for inclusion in stage 2, which was based upon a rank ordering of institutions on a set of quantitative learning and teaching performance indicators. In 2008, the performance indicators for the scheme were: Student Progress Rate, Student Retention Rate, Graduate Full-Time Employment, Graduate Further Full-Time and Part-Time Study, Graduate Satisfaction with Generic Skills, Graduate Satisfaction with Good Teaching and Graduate Overall Satisfaction. A sophisticated statistical process for adjusting these indicators to take account of institutional variation is outlined in the paper. The adjustment factors included: Mix of disciplines, Admission basis criteria, Type of attendance, Mode of attendance, Indigenous status, Gender, Disability, Non-English Speaking Background, Age, Location, Socio-Economic Status, Unemployment rate, Number of students, Level of study and New to higher education. Within each broad field of study, institutions are placed in bands based upon the combination of indicators, and funding is allocated to institutions. Even with the availability of a standardised, evidence-based, effectively compulsory, student experience survey (Course Experience Questionnaire) the establishment and maintenance of a national teaching comparator has proved difficult and controversial. The scheme is not being continued from 2010.

9.7 High-Quality Learning Outcomes

Given the problems with surveys of student experience, it is of little surprise that the focus is shifting to establishing common national and international ground in the outcomes of student learning. Since learning is the main reason why teaching is carried out, an assessment of the quality of learning is also a valid component of ranking systems. But similar quality questions remain – what is high quality learning? In a major UK study on student engagement and high quality learning, the Enhancing Teaching-Learning (ETL) Environments research team describes desired student outcomes more in terms of ways of thinking and practising in the discipline or professional area (Hounsell et al. 2005). Ways of thinking and practising ‘capture the richness, depth and breadth of what the students could learn through engagement with a given discipline or subject area. It rests on a secure foundation of subject knowledge and understanding and could also encompass

subject-related skills, conventions and practices for communicating within the subject, and salient values and attitudes' (p. 5).

To achieve what Hounsell and colleagues suggest is likely to be extremely difficult, but the struggle has begun. Marginson and van der Wende (2007) describe the developments as follows: 'Based on these trends in relation to the job market, there are moves to apply the terminology for normal education to higher education. One example is the standardisation of higher education in curriculum, testing, credit hours and degree awarding requirements. Through the Bologna Process, European countries are moving towards a common structure for bachelors, master and doctoral degrees, with similar achievement criteria and credit hours systems. In some countries (e.g. the UK), benchmarks for each discipline area have been set within quality assurance systems. Based on the notions of standardisation, the OECD is developing standardised tests for college graduates similar to those for secondary school students (the so-called 'PISA'). If standardisation is widely applied, the shape of higher education will be totally different from what it is now. Curricula will be standardised in each discipline area, standardised tests will be given to most college students, each course will carry the same credit hours, and degree requirements will be similar across national borders' (p. 319).

Similar developments are happening in the USA. For example, the Middle States Commission on Higher Education has a long-standing commitment to outcomes assessment and to student learning outcomes in particular. In their publication *Characteristics of Excellence in Higher Education* (Middle States Commission on Higher Education 2006), excellence is defined as meeting 14 accreditation standards. Seven of the standards are related to educational effectiveness (Admissions and Retention; Student Support Services; Faculty; Educational Offerings; General Education; Related Educational Activities (Basic Skills; Certificate Programmes; Experiential Learning; Non-Credit Offerings; Branch Campuses, Additional Locations, and Other Instructional Sites; Distance or Distributed Learning; Contractual Relationships and Affiliated Providers) and Assessment of Student Learning). Each of the standards is defined in terms of outcomes. For example, for Student Admissions and Retention outcomes are: Statements of expected student learning outcomes and information on institution-wide assessment results, as appropriate to the programme offered, available to prospective students (Prosser M., 2009, Personal communication).

But here too, there will be difficulty as the focus is likely to be on achievement of a minimum or threshold level, and such an approach cannot therefore be used to assess the degree of 'value added'.

9.8 Conclusions

The analysis described in this chapter does not provide any clear guidance for national ranking systems based on university teaching that could be of use to inform potential students, or the development of global indicators of teaching

quality. Part of the problem could be that the search for the answers is focused mainly on the easier options rather than looking at what might be needed. Even a cursory glance at the way the research indices are measured (and accepted) reveals the enormous amount of international peer review resource that is expended in the process. It is highly likely that the same international peer review framework is needed for teaching. At a national level, this has been growing in some countries for several decades, for example in the allocation of national teaching grants in the 1990s (Australian Committee for the Advancement of University Teaching and the Swedish Council for the Renewal of Undergraduate Education) and in teaching awards (e.g. National Teaching Fellowship Scheme (Higher Education Academy 2010b)). The scholarship of teaching movement (ISSOTL 2010) is also encouraging international peer review of scholarly teaching artefacts. Despite this progress, descriptions of international standards and universal acceptance of them are still many years away.

If and when international teaching/learning standards are developed, their use in rankings may still be seen to be undesirable. As Teichler (2008) has commented, ranking reports guide vertical classification of HEIs that is likely to have adverse effects on institutional diversity. In addition, ranking indicators are overly simplified and do not necessarily reflect the quality of education. Instead, international ranking surveys (e.g. Shanghai Jiao Tong's Ranking) put major emphasis on research performance instead of teaching.

From the perspective of preparing acceptable global ranking systems that include a prominent place for teaching, there are clearly many obstacles to overcome, but from the point of view of enhancing the quality of student learning, attempts at ranking systems may not be such a bad idea. 'The savviest universities are using the rankings phenomenon to catalyse and establish keener performance evaluation internally at a faculty, department and individual staff member level. Driving it down to this level can help build actionable metrics as opposed to abstract statistics and this can lead to a university being able to revolutionise their performance in education and research, and in time, as a side-effect rather than an objective, improve their performance in rankings' (Sowter 2009).

An example of what Sowter has described exists at the University of Sydney. Nationally, the Course Experience Questionnaire is used in Australia to capture the comparative experience of university graduates. The release of data, at times as much as 6 years after students have begun their first year of study, is not sufficiently current to support change (some courses will have gone through one and sometimes two revisions in that period). To facilitate the improvement process, the University of Sydney introduced a Student Course Experience Questionnaire (SCEQ) that is completed by students in all years of the programme on their experience of their course to that point. This provides the 'bigger picture' data that is up to date, and is a measure of the quality of the course at that time. The SCEQ includes the same scales as the CEQ (Good Teaching and Generic Skills) but also scales on Clear Goals and Standards, Appropriate Assessment and Appropriate Workload. To enable change to be addressed at the level of the individual teacher, a third level of questionnaire (Unit of Study Evaluation, USE) was designed with 12 questions

including one on each of the five scales of the SCEQ. As such, an individual teacher has a 'direct line of sight' from the USE to the CEQ. Relations of this sort are an essential criterion of any national indicator of university teaching.

There is additional evidence that teaching indicators can generate positive change. The Institute for Higher Education Policy notes that the institutions included in their case studies '… continue to point to their changing practices that alter input indicators – increasing selectivity, favouring research over teaching, and strengthening the faculty profile – [while] a number of institutions are also reporting changes to practices directly related to student learning and success. Institutions that use their rankings to prompt change in areas that directly improve student learning experiences demonstrate that rankings can lead to positive change in teaching and learning practices' (Institute for Higher Education Policy 2009: 3).

References

Altbach, P. (2006). The dilemmas of ranking. *International Higher Education, 42*, 2–3.

Argyris, C., & Schön, D. (1974). *Theory in practice: Increasing professional effectiveness*. San Francisco: Jossey Bass.

Astin, A., & Chang, M. J. (1995). Colleges that emphasize research and teaching. *Change, 27*(5), 44–49.

Australian Council for Education Research. (2010). Australasian Survey of Student Engagement. Retrieved June 23, 2010, from http://www.acer.edu.au/documents/AUSSE_Brochure.pdf

Benowski, K. (1991). Restoring the pillars of higher education. *Quality Progress, 24*(10), 37–42.

Biggs, J. (2001). The reflective institution: Assuring and enhancing the quality of teaching and learning. *Higher Education, 41*(3), 221–238.

Boyer, E. (1990). *Scholarship reconsidered: Priorities of the professoriate*. Princeton: Carnegie Foundation for the Scholarship of Teaching.

Brooks, R. L. (2005). Measuring university quality. *The Review of Higher Education, 29*(1), 1–21.

Clarke, M. (2005). Quality assessment lessons from Australia and New Zealand. *Higher Education in Europe, 30*, 183–197.

Department of Education, Science and Training, Australian Government. (2004). *Learning and teaching performance fund: Issues paper*. Retrieved June 23, 2010, from http://www.dest.gov.au/NR/rdonlyres/8F857836-F4AE-4D74-952E-3906FF5807B4/10254/ltpf_issues_2004.pdf

De Miguel, J. M., Vaquera, E., & Sanchez, J. D. (2005). Spanish universities and the ranking 2005 initiative. *Higher Education in Europe, 30*, 199–215.

Dill, D. D., & Soo, M. (2005). Academic quality, league tables, and public policy: A cross-national analysis of university ranking systems. *Higher Education, 49*, 495–533.

Eccles, C. (2002). The use of university rankings in the United Kingdom. *Higher Education in Europe, 27*, 423–432.

Federkeil, G. (2002). Some aspects of ranking methodology – The CHE-ranking of German universities. *Higher Education in Europe, 27*, 389–397.

Filinov, N. B., & Ruchkina, S. (2002). The ranking of higher education institutions in Russia: Some methodological problems. *Higher Education in Europe, 27*, 407–421.

Gibbs, G. (1995). The relationship between quality in research and quality in teaching. *Quality in Higher Education, 1*(2), 147–157.

Glassick, C. E., Huber, M. T., & Maeroff, G. I. (1997). *Scholarship assessed: Evaluation of the professoriate*. San Francisco: Jossey-Bass.

Guarino, C., Ridgeway, G., Chun, M., & Buddin, R. (2005). Latent variable analysis: A new approach to university ranking. *Higher Education in Europe, 30*(2), 147–165.

Hattie, J., & Marsh, H. W. (1996). The relationship between research and teaching: A meta-analysis. *Review of Educational Research, 66*, 507–542.
Hau, H. (1996). Teaching quality improvement by quality improvement in teaching. *Quality Engineering, 9*(1), 77–94.
Harvey, L., Burrows, A., & Green, D. (1992). *Criteria of quality in higher education report of the QHE project*. Birmingham: The University of Central England.
Harvey, L., & Green, D. (1993). Defining quality. *Assessment and Evaluation in Higher Education, 18*, 8–35.
Higher Education Academy. (2010a). *National student survey*. Retrieved June 23, 2010, from http://www.hefce.ac.uk/learning/nss/
Higher Education Academy. (2010b). *National teaching fellowship scheme projects*. Retrieved June 23, 2010, from http://www.heacademy.ac.uk/ourwork/professional/ntfs/projects
Hounsell, D., Entwistle, N., Anderson, C., Bromage, A., Day, K., Hounsell, J., Land, R., Litjens, J., McCune, V., Meyer, E., Reimann, N., & Xu, R. (2005). Enhancing teaching-learning environments. *Final report to the Economic and Social Research Council on TLRP project L13925109*. Retrieved June 23, 2010, from http://www.etl.tla.ed.ac.uk/docs/ETLfinalreport.pdf
Institute for Higher Education Policy. (2009). *Impact of college rankings on institutional decision making: Four country case studies*. Retrieved June 23, 2010, from http://www.ihep.org/assets/files/publications/g-l/ImpactofCollegeRankings.pdf
ISSOTL. (2010). *International Society for the Scholarship of Teaching and Learning*. Retrieved 23 June, 2010, from http://www.issotl.org/
Kuh, G. D. (2001). *The national survey of student engagement: Conceptual framework and overview of psychometric properties*. Bloomington: Indiana Center for Postsecondary Research and Planning.
Ledesma, J. R. (2002). Latin America: A different reality. *Higher Education in Europe, 27*, 467–474.
Liu, N., & Cheng, Y. (2005). The academic ranking of world universities. *Higher Education in Europe, 30*(2), 127–136.
Liu, N., & Liu, L. (2005). University rankings in China. *Higher Education in Europe, 30*, 217–227.
Lizzio, A., Wilson, K., & Simons, R. (2002). University students' perceptions of the learning environment and academic outcomes: Implications for theory and practice. *Studies in Higher Education, 27*, 27–52.
Marsh, H. W., & Cheng, J. H. S. (2008). *National student survey of teaching in UK universities: Dimensionality, multilevel structure, and differentiation at the level of university and discipline: Preliminary results*. Retrieved June 23, 2010, from http://hlst.ltsn.ac.uk/assets/York/documents/ourwork/research/surveys/nss/NSS_herb_marsh-28.08.08.pdf
Marsh, H. W., & Hattie, J. (2002). The relation between research productivity and teaching effectiveness: Complementary, antagonistic, or independent constructs? *The Journal of Higher Education, 73*(3), 603–641.
Marginson, S., & van der Wende, M. (2007). To rank or to be ranked? The impact of global rankings in higher education. *Journal of Studies in International Education, 11*(3), 306–339.
Middle States Commission on Higher Education. (2006). *Characteristics of excellence in higher education: Eligibility requirements and standards for accreditation*. Retrieved June 23, 2010, from http://www.msche.org/publications/CHX06060320124919.pdf
Morse, R., & Flemingan, S. (2009). *How we calculate the college rankings*. Retrieved June 23, from http://www.usnews.com/articles/education/best-colleges/2009/08/19/how-we-calculate-the-college-rankings.html?PageNr=1
National Survey of Student Engagement. (2010a). *About the national survey of student engagement*. Retrieved June 23, 2010, from http://nsse.iub.edu/html/about.cfm
National Survey of Student Engagement. (2010b). *Benchmarks of effective educational practice*. Retrieved June 23, 2010, from http://nsse.iub.edu/pdf/nsse_benchmarks.pdf
Patrick, W., & Stanley, E. (1998). Teaching and research quality indicators and the shaping of higher education. *Research in Higher Education, 39*(1), 19–41.

Perspektywy, W. S. (2002). Ten years of rankings. *Higher Education in Europe, 27*, 399–406.
Prosser, M., & Trigwell, K. (1999). *Understanding learning and teaching: The experience in higher education*. Buckingham: Open University Press.
Ramsden, P. (1991). A performance indicator of teaching quality in higher education: The Course Experience Questionnaire. *Studies in Higher Education, 16*, 129–150.
Ramsden, P., Margetson, D., Martin, E., & Clarke, S. (1995). *Recognising and rewarding good teaching in Australian higher education*. Canberra: Committee for the Advancement of University Teaching, Australian Government Publishing Service.
Shulman, L. S. (1987). Assessment for teaching: An initiative for the profession. *Phi Delta Kappan, 69*(1), 38–44.
Shulman, L. S. (1988). A union of insufficiencies: Strategies for teacher assessment in a period of educational reform. *Educational Leadership, 46*(3), 36–41.
Sowter, B. (2009). *University rankings: There can be no "right answer"*. Retrieved June 23, 2010, from http://qsiu.wordpress.com/2009/04/24/university-rankings-there-can-be-no-right-answer/
Stromquist, N. P. (2007). Internationalization as a response to globalization: Radical shifts in university environments. *Higher Education, 53*, 81–105.
Surridge, P. (2008). The National Student Survey 2005–2007: findings and trends. *A report to the Higher Education Funding Council for England*. Retrieved June 23, 2010, from http://www.hefce.ac.uk/pubs/rdreports/2008/rd12_08/
Tam, M. (2001). Measuring quality and performance in higher education. *Quality in Education, 7*(1), 4–54.
Taylor, M. (2003). Teaching capabilities and professional development and qualifications framework project: stage one. Unpublished Report, Melbourne: RMIT University
Teichler, U. (2008). Diversification? Trends and explanations of the shape and size of higher education. *Higher Education, 56*(3), 349–379.
The Princeton Review. (2010). *Best value colleges methodology*. Retrieved June 23, 2010, from http://www.princetonreview.com/bvc-methodology.aspx
The Times Higher Education (2010). Methodology: A simple overview. Retrieved June 23, 2010, from http://www.topuniversities.com/university-rankings/world-university-rankings/methodology/simple-overview
Trigwell, K., Martin, E., Benjamin, J., & Prosser, M. (2000). Scholarship of teaching: A model. *Higher Education Research and Development, 19*, 155–168.
Yair, G. (2008). Can we administer the scholarship of teaching? Lessons from outstanding professors in higher education. *Higher Education, 55*(4), 447–459.
Yonezawa, A., Nakatsui, I., & Kobayashi, T. (2002). University rankings in Japan. *Higher Education in Europe, 27*, 373–382.

Part III
Social Impacts of University Rankings

Chapter 10
On the Efficacy of Raising Your University's Rankings

Christopher C. Morphew and Christopher Swanson

10.1 Introduction

Rankings are the Swiss knife of higher education – they are a single tool with many uses. Like many other universities, Texas Tech University utilizes rankings as a barometer to judge whether the university exhibits dimensions of quality. (The term "universities" will be used to describe all postsecondary institutions throughout this chapter.) The "Goal Two: Academic Excellence" section of its 2005 strategic plan cites rankings 12 times. Three of the nine objectives in this section of the plan are explicitly aimed at improving the institution's national ranking, whether it be in selectivity, grants, scholarly productivity, or the quality of the university's library system (Texas Tech University 2005). The use of rankings as a measure of a college or university's excellence, improvement in quality, prestige, character, hipness, or value is ubiquitous. The pervasiveness of ranking systems has spread to institutions outside the United States as well. At world-renowned institutions like the University of Melbourne in Australia, for example, international rank is so important it occupies the second highlight on the "About the University" page, sandwiched between the institution's foundation date and the number of enrolled students (University of Melbourne 2010). Even lesser-known institutions, like the University of Kwazulu-natal in South Africa use higher education rankings in creating strategic plans as well as guideposts in determining institutional quality (University of Kwaxulu-natal and Strategic 2007). As these examples demonstrate, universities have adopted the use of rankings as a means of assuring internal actors that the institution is on course toward its goals.

Other institutions use rankings as a signal flare, to highlight their quality for external constituents. Benedictine College, in Atchison, Kansas, for example, proudly displays the *US News* "Best Colleges" emblem on its homepage and notes its top 20 ranking in the *Newman Guide to Choosing A Catholic College* on its "About Benedictine" website (Benedictine 2009). Rankings assure prospective students

C.C. Morphew (✉) and C. Swanson
Department of Educational Policy and Leadership Studies, University of Iowa, Iowa City, IA, USA
e-mail: christopher-morphew@uiowa.edu

J.C. Shin et al. (eds.), *University Rankings*, The Changing Academy – The Changing Academic Profession in International Comparative Perspective 3,
DOI 10.1007/978-94-007-1116-7_10,

and their parents that Benedictine is a legitimate Catholic institution of quality. The use of rankings as a communicative tool is so powerful and widespread that it has pandemically spread beyond the institutional level. It has become common for a college or university – and its academic units – to explicitly cite popular rankings, such as *US News* or *Business Week,* as arbiters of its caliber. This is true of privates and publics, elites and non-elites. The College of Engineering at the University of California, Berkeley, acknowledged in many national and international rankings as an eminent postsecondary institution, cites its position in national rankings, particularly its specific position relative to archrival Stanford University, on its webpage (UC Berkeley 2009). Caltech, also a renowned US university, devotes space on its webpage to documenting the university's rankings in international rankings, such as *The Times Higher Education Supplement* (Caltech 2009). Universities use rankings, in their current form, to provide both informational and promotional properties to internal and external constituents alike.

This chapter provides an analysis of (a) how universities are controlled by higher education rankings; (b) how universities react to rankings; (c) the importance of reputation – a major factor in the rankings – as an intangible resource; (d) equity concerns relevant to higher education rankings; and (e) resultant lessons regarding the efficacy of pursuing a change in rankings. The goal is not to arrive at a normative conclusion, but to broadly assess how a university's leaders might utilize information about higher education rankings to make relevant institutional policy. Further, this chapter focuses primarily on colleges and universities in the United States, but its topics can be broadly applied to international institutions.

10.2 Higher Education as Fertile Ground for Rankings

It should come as no surprise that rankings are a popular device for universities. Because identifying quality is so difficult in institutionalized fields such as higher education, organizational myths of structure and behavior take on important meaning (Meyer and Rowan 1977). In technical fields, for example, there is less need for rankings that incorporate "soft" variables such as reputation and the relationship between inputs and outputs. Instead, organizations in these fields can be compared on objective measures explicitly linked to the quality and quantity of their outputs. However, these measures are unavailable in higher education and, in lieu of these, rankings are simultaneously myth making and sense making.

There are at least two reasons to explain why rankings have become so valuable, both of which relate to the institutional nature of the higher education environment. First, because both the products and technology of higher education are nebulous and hard to measure, rankings provide a seemingly objective input into any discussion or assessment of what constitutes quality in higher education. As Gioia and Thomas (1996) point out, there are "few bottom-line measures like profit or return on investment that apply to" higher education (p. 370). To fill this void, organizations inside and outside higher education have created ranking systems.

Higher education rankings also take advantage of the fact that the institutionalized nature of the industry makes mimicry more useful and likely. Because it is so difficult – for insiders and outsiders – to judge the quality of a university's technology and outputs, it is often more cost-effective and useful for institutions to copy the structures and behaviors of universities perceived as being successful. The ambiguity of organizational goals within higher education makes this kind of modeling convenient and predictable (DiMaggio and Powell 1991). Programs such as the German Excellence Initiative, designed to thwart the dominance of international rankings by the USA and U.K. universities, illustrate the universality of this type of mimicry in higher education (Labi 2010).

Rankings produced by organizations such as *US News* or *The Times* complement and buttress the already isomorphic nature of higher education. By codifying and ordering the successful practices and structures of elite organizations such as Harvard or Oxford, rankings produce a navigable roadmap for less-prestigious institutions to follow. Rankings utilize the key attributes of elite institutions. These key attributes are then weighed heavily in the rankings algorithm, which produce the assumptive results. Naturally examined, dissected, and ultimately mirrored by their non-elite counterparts, elite universities establish the standard by which all institutions are gauged. By nuancing the differences between the seemingly successful and the seemingly less-successful institutions, the creation of a set of rankings inevitably quantifies the various academic dimensions of all institutions. This quantification of relationships between institutions exacerbates and amplifies the mimetic tendencies already found in higher education. While mimicry often occurs without the existence of rankings, they further legitimate practices by substituting improvement in rankings for evidence of real improvement. Devinney et al. (2008) argues that the "dark side" of mimetic isomorphism in higher education is that institutions will stop experimenting and instead favor herd behavior that is ultimately destructive to their organizational field. In short, it is predictable and problematic that rankings catalyze the mimetic tendencies of organizational behavior in higher education.

10.3 The Control Exerted by Higher Education Rankings

Conceptually, the purpose of publicizing rankings and tightly coupling strategic actions to rankings can be explained. In this section, we investigate what the limited literature on the subject tells us about how and why universities utilize rankings.

Researchers who studied the reaction of university leaders when introduced to rankings agree that, initially, administrators viewed rankings such as *US News* as less than legitimate. The Dean of Harvard's Law School referred to the 1998 rankings as "Mickey Mouse," when asked about their relevance to the field (Parloff 1998). A decade or so later, even given this derision, rankings have become so legitimate as to influence the behavior and culture of law schools. Sauder and Espeland (2009) describe the influence of rankings as impossible to ignore and

difficult to manage. Because they have come to occupy a central position in the application process for prospective law school students, rankings have come to play a permanent, indelible role for the schools. The mark they receive from *US News* is a kind of tattoo that instantly and powerfully communicates their standing in the larger field. Sauder and Espeland use Foucalt's conception of discipline to make sense of how law schools are forced to internalize and incorporate the values of rankings. However, they could just as easily be describing Weber's (1958) concept of the iron cage, which focused on the means by which organizations and their actors are increasingly constrained by a bounded rationality predicated on goals. In this light, the goal of being a "quality institution" has forced and legitimated the use of rankings onto law schools, as well as other institutions. One of the more explicit pieces of evidence that can be found to substantiate this argument is found in Loyola University's (Louisiana 2009) strategic plan, which states:

> To enhance our reputation and stature, as reflected in the rankings of *U.S. News and World Report*, we are committed to a university-wide rethinking of our programs in a way that builds upon our strengths and utilizes new initiatives that respond to national needs and student demands. Such an approach seeks to increase demand and attract more and better students, which will decrease the need to discount tuition, while allowing Loyola to attract students from deserving communities and shape our incoming classes. An increase in ranking will directly affect an increase in revenue.

Rankings have been so successful in demarcating what constitutes quality in higher education that university strategic plans now commonly refer to them as a valid arbiter of quality. Why? The discussion above notes the institutional nature of higher education, but there is evidence that other organizational types respond aggressively to being ranked, particularly when that rank threatens their legitimacy within a specific organizational field. Chatterji and Toffel's (2008) research on the effects of third party environmental ratings on for-profit firms, for example, delineates how firms with low environmental ratings responded positively to such a less-than-favorable rating. Further, firms with lower environmental ratings improved their performance on these criteria, as compared with those rated higher. Both institutional and strategic choice theories explain these behaviors. Organizations facing the prospect of being delegitimized by a third-party rating must choose how to respond, particularly if that rating carries credence by important constituents. The research suggests that firms with particularly low ratings are more likely than their higher-rated peers to respond with practices that leverage the "low hanging fruit" available to them and thus improve their rating. These finding suggests that ratings should incorporate both "sticks" and "carrots," in order to affect change in high- and low-rated organizations and avoid the negative convergence that may accompany ratings that focus on problems only.

In a similar fashion, university rankings have determined, even codified, what types of organizational behaviors and practices are legitimate (Wedlin 2007; Hazelkorn 2009; Chap. 11). This is particularly true in the case of law and business school rankings, where research suggests that students – and, as a result, deans – have come to view rankings as "the" primary determinant in choosing to apply to and/or attend a specific university (Elsbach and Kramer 1996). The results are

predictable; organizations, regardless of their status, conform to the rankings agenda, even as new rankings are introduced by those who, for example, fear their organizational or national identity is being marginalized. Wedlin's (2007) work on the compelling nature of the international MBA programs suggests that faculty and administrative staff at business schools are seeing their exclusive role in shaping curricular and programmatic decisions usurped by rankings that prescribe what "a good and proper international business school is, or what it should be; what programs and features are important, how schools should structure and carry out work" (p. 28). The work of others (Sauder and Espeland 2009; Hazelkorn 2007; Espeland and Sauder 2007; Martins 2005) substantiates these claims.

10.4 Reacting to Rankings

In response to the public's and higher education's demonstrated embrace of rankings, universities are adjusting their educational practices and strategies to obtain a favorable rank from both the media organization and their consumers. Evidence showcasing the beneficial outcomes associated with rankings is relatively young; yet, there are some intriguing conclusions deserving further analysis by researchers and policymakers. From a macro point of view, the research suggests that universities have relatively little control over their rankings, whereas, from a micro perspective, smaller, yet important changes may be possible as a function of concentrated changed behavior. Overall, a paradox is emerging: Rankings are a game everyone plays, but a game with constantly shifting rules that no one can control.

Several findings have been confirmed and reconfirmed in multiple studies on the effects of rankings. For example, a higher rank in a given year, controlling for other factors, will result in more applicants for admission, a lower admissions yield (higher selectivity), and higher median test scores among both the applicants and the enrolled student pools (Monks and Ehrenberg 1999). Prospective applicants notice and respond to rankings. Other studies that examined similar phenomena (Bowman and Bastedo 2009; Meredith 2004; Sauder and Lancaster 2006) suggest that ranking outcomes associated with student admissions compound over time. That is, an improvement in rank in 1 year, in turn, creates a favorable situation for the institution in subsequent years. However, this phenomenon cuts both ways, in that lower rankings year after year will produce subsequent applicant and enrolled student populations that exacerbate the inability of the university in questions to attract and enroll high ability students.

Beyond admissions, there are other institutional outcomes researchers have linked to rankings. Bastedo and Bowman (2009), for example, found rankings to directly affect the funding of research and development from government, industry, and foundations, as well as the total amount of alumni donations. This effect confirms the previous assumptions of financial contributions to higher education in that donors utilize rankings to be associated with successful universities. Current and past research on the subject documents how donors – alumni and those without

connections to the university in question – are more likely to contribute when tangible indicators of success are present, including but not limited to a growing endowment or a successful athletics program (Leslie and Ramey 1988; Ehrenberg and Smith 2003; McCormick and Tinsley 1990). It is apparent that rankings work like other signaling devices in higher education. Better students, faculty, and wealthy donors are attracted to those universities perceived as better, more prestigious, or higher quality because of the perceived benefits of being associated with these successful organizations.

Due to these financial and non-financial benefits, institutions eagerly find ways to improve their rankings. Because certain ranking schemes take into effect more easily manipulated data, universities employ a number of gaming techniques to improve their position in the rankings. For example, a university may ignore adjunct instructors altogether when reporting the percentage of full-time faculty employed – a known function of many rankings. In the most recent round of *US News* rankings, several well-known public universities including the Georgia Institute of Technology, the Pennsylvania State University, the University of Iowa, North Carolina State University, and the University of Nebraska reported faculty data without including many or all adjuncts, despite the magazine's explicit request for institutions to include adjuncts in their self-reported calculations (Jaschik 2009). Several explanations were given. Adjuncts were considered employees, not faculty, at Penn State. North Carolina State and Iowa considered adjuncts faculty only if they held permanent appointments, which most did not. In any case, these universities' ranking benefited from this misreporting, which can be reasonably surmised as the intent.

The conjuring of numbers is only one of many schemes used by institutions trying to improve their rank. Colleges and universities employ a number of other manipulative tricks as well, all born from and focused onto the various components used in the rankings algorithm. For example, to manipulate the "beginning characteristics" in *US News*, a 15% component of the overall index score, institutions have been found to intentionally misreport admissions data as well as encourage unqualified students to apply, only to coldly reject them later – boosting the selectivity rate (Ehrenberg 2002). Other institutions misreport a current student's single gift as a multi-year gift, enabling the institution to claim these donations as alumni gifts (Golden 2007). Law schools spend over $100,000 a year in creating, printing, and sending glossy marketing brochures to other law school administrators hoping to influence "reputation" scores, a 25% component (Espeland and Sauder 2007). While these gaming techniques seem underhanded and unrelated to institutional quality, they serve as a means to a more favorable ranking end. If anything, "gaming challenges the legitimacy of rankings by subverting their appearance as accurate representations of the schools they measure […] but gaming simultaneously reinforces the legitimacy of rankings by furthering educators' investment in them" (Sauder and Espeland 2009: 78). Stated otherwise, the gaming techniques practiced by contemporary institutions of higher education reveal both the destructive and staying power of rankings.

Despite a strong desire to improve in the rankings, the amount of control institutions have in the process is highly debatable. One study finds around 70–80% of the

variability between annual rankings is transitory "noise" and disappears within 2 years. These results suggest that rankings do very little to document or reward real improvements in quality (Dichev 2001). Similarly, the monolithic nature of being an "elite" institution is impressive. For example, in *US News,* the dominant United States ranking guide, only 29 schools occupied the top 25 spots between 1988 and 1998, and 20 institutions never fell out of the top 25. In reality, it is nearly impossible for any university outside the top 25 to break into this elite group, and aspirations to do so represent, in the vast majority of cases, organizational daydreaming. Moreover, the fierce competition for a top spot among all institutions, in the zero-sum game of rankings, only serves to make positive movement that much harder. In a recent survey of higher education administrators, Hazelkorn (2007) noted that 93% and 82% of respondents wanted to improve their national and international rank, respectively. Additionally, she found that 70% wanted to be in the top 10% nationally and 71% in the top 25% internationally. Devinney et al. (2008) take the impossibility of institutional control one step further in providing evidence that "most of the critical attributes that matter to the rankings are correlated with structural factors" (p. 10), or factors that are either impossible, or financially impractical for institutions to manipulate. Evidence mounts of the paradox of pursuing a higher ranking: An increasing desire to improve rank often belies the decreasing ability to do so.

10.5 The Power of Reputation

Reputation is an intangible organizational asset that is both hard to construct and, if lost, hard to recover. The empirical evidence on the subject indicates that organizations, including universities, are right to worry about their reputation and its attached benefits. Studies of for-profit firms have demonstrated that managers value an organization's reputation as *the* most important intangible resource a business can have, more important than, for example, employee know-how. However, without the technical data to delineate organizational strengths and weaknesses often found in for-profit enterprises, reputation is likely more important in fields like higher education. Here though, the reliance upon reputation can drastically exacerbate its effect on internal and external constituents. Regardless of industry, an organization's reputation is complemented by the fact that this resource is very difficult to develop and requires a long period to rebuild (Hall 1992; Deephouse 2000; Rindova et al. 2005).

Widely cited by managers as critically influential, reputations are the invisible, unquantifiable "dark matter" of organizations. Difficult to see or manipulate, very few studies have calculated the exact impact of reputation on organizational performance. However, those that have project a unified voice: Reputation has "considerable significance with respect to the sustainability of advantage" (Hall 1992: 143). Attributing various performance measures, like financial success, solely to an organization's reputation is exacting at best, but evidence suggests reputation can significantly affect performance. In an analysis of banks in the Twin Cities (Minnesota, USA) area, one study concluded that a bank with a relatively

strong reputation enjoys a significant financial advantage in competition with other banks. This advantage manifests itself in several important outputs, including lower costs, the ability to price goods and services at a premium, and a competitive advantage that is hard-to-overcome (Deephouse 2000). Similar advantages can be found in higher education where universities sporting strong reputations relative to their peer group can raise tuition price and enjoy increased numbers of applicants and revenues (Weisbrod et al. 2008). Conversely, universities without such strong reputations may be forced to cut the cost of tuition, in order to attract greater numbers of students, who would otherwise apply to similarly priced universities with better reputations (Jaschik 2008). Broader conclusions of the value of reputation are supported by Roberts and Dowling (2002), who argued in their 14-year analysis of *Fortune 1000*'s *America's Most Admired Corporations* that reputation served as a buttress for better long-term financial performance. From this perspective, although reputation is usually considered an untechnical or "soft" criterion, it is actually a kind of "hard" asset that won't erode over time and can serve an organization during periods of stability and instability. These findings suggest that reputation – while invisible and difficult to control – is critical to isolating the top performers from the rest of the field.

Reputations hold value precisely because of the competitive advantage they provide and the relative costs and/or ability associated with procuring a similar positive reputation. Organizations may use other means of substituting for a positive reputation, such as guarantees or warranties, but these substitutes have real costs and may not provide similar value for the organization or the consumer (Klein et al. 1978. In higher education, the lack of a positive reputation can limit the approaches available to universities in their marketing to students. Metropolitan State College in Denver, Colorado, which lacks a particularly strong reputation (Tier 4 among *US News* liberal arts colleges) recently made headlines by offering free remediation to any of its teacher education graduates who were unsuccessful in the classroom (*Denver Post* 2009). Similarly, Doane College, a lower-ranked (Tier 3 among *US News* liberal arts colleges) small baccalaureate college in Doane, Nebraska guarantees a 4-year graduation to all full-time students. If not, students receive any additional courses tuition-free (Doane College 2009). Even relatively highly-ranked Juniata College (#85 in *US News* liberal arts colleges) offers its students a "buy four, get one free" guarantee, providing a free year of tuition to all full-time students who fail to graduate in 4 years or less (Weggel 2007). The highest ranked universities need not offer such warranties, potentially saving them money.

The difficulty in higher education – and other organizational fields – is that reputation is a resource that cannot be easily purchased or improved. Positive organizational reputations may be the product of historical incidents that cannot be replicated, making them "imperfectly imitable" (Barney 1991: 115). Similar to the monolithic nature of rankings, organizations with positive reputations find it relatively easy to maintain them, while those with less-positive reputations find it very difficult to improve their reputation, particularly relative to organizations in the same field with longstanding positive reputations. On the other hand, recent studies of *US News* rankings of US universities show how a move in ranking, particularly

when a university changes tiers, can have a positive impact on the future peer assessments of the same university. In other words, while reputation is difficult to improve, it is not impossible, especially when reflected in the rankings, because peer assessment of an institution can be changed over time through improvements in selectivity and the utilization of resources (Bastedo and Bowman 2009; Bowman and Bastedo 2009). This finding is at odds with decades of reputational stability in American universities:

> Reputational surveys of American universities conducted in the 1950s, 1960s, and 1970s revealed an academic pecking order of remarkable stability. In the competition for top-twenty rankings, rarely was there a new institutional face (Graham and Diamond 1997: 2)

The good news for those in higher education with strong positive reputations (and bad news for the rest) is that reputation carries tremendous weight in many national and international rankings. For example, rankings by *AsiaWeek,* Education18.com, Melbourne, *The Times* World University Rankings, Netbig, *US News*, Wuhan, and *MacLeans* include a variable linked to reputation. Among these, three (Education18.com, *The Times* and *US News)* weight reputational scores very heavily – at least 25% (Usher and Savino 2006). The use of reputation as a variable will make it nearly certain that these rankings will display relatively little variation in their top-rated universities.

10.6 Equity Concerns and Students' Use of Rankings

Research on college choice depicts a dynamic process, whereby the decision to apply to a specific university is a function of both self-selection and societal context. Generally described as a three-stage process, students first consider their options while assuming information about what kind of university they want to attend, and self-assess their probability in attending such an institution. The assumptions held by students are largely a function of socioeconomic status (SES) (Hossler et al. 1999). Using these initial constraints, students develop a "choice set" of universities, often excluding those viewed as unaffordable. This set may contain, however, "safety schools" as well as "reach" or aspirational choices, based on selectivity and the cost of the university (Hossler and Gallagher 1987).

The research on the outcomes of this pre-application selection stage is quite clear: SES plays a substantial role in the college choice process. Lower-income students, constrained by their socioeconomic status, are inevitably less likely to choose a selective, more expensive institution than their more privileged peers (Steinberg et al. 2009). Lower-income students are also more likely than their peers, controlling for other factors, to choose to attend a university close to home (McDonough 1997; St John et al. 2001; Pryor, et al 2008).

SES also determines, to some degree, the type of information prospective students use throughout the college choice process. Because educational quality proves difficult to assess, students tend to utilize the admissions selectivity indicators as a means of gleaning the differences between institutions. From here, the vast majority

of students engage in "self-selection," the process of applying only to institutions in which they can both gain admission as well as afford (Hossler and Litten 1993; Hossler et al. 1999; McDonough 1997); thus, the students' ability to accurately judge a university's admission standards is extremely crucial. Knowledge about universities, however, is not evenly distributed among students. Students from underrepresented backgrounds often have less access to informational resources such as high school counselors, who may have little time to invest in shaping students' postsecondary aspirations (McDonough 1997; McDonough and Calderone 2006). In short, students from lower socioeconomic backgrounds, relative to their peers, choose among less-prestigious, lower-ranked institutions and have less access to critical information.

As a substitute for this institutional knowledge, rankings are often sold as a means to "find the best college for you" and a tool to "find your perfect fit" (US News & World Report 2010). While this function is frequently debated among practitioners and non-practitioners alike, nonetheless, who actually uses the rankings and for what purposes becomes a very important variable in the college choice process. In fact, the utilization of rankings is strongly correlated with student's socioeconomic status. Students from families with higher levels of income and education use rankings more often and are more likely to report university rank as an important factor in their college choice decision as compared to poorer students who use rankings less often and find a university rank not at all important. Examining the college choice process and the role rankings play, McDonough et al. (1998) argued that, instead of aiding in finding a college that "fits" a student, rakings are used by high-income students to signal their status and are "merely reinforcing and legitimizing these students' status obsession" (p. 531).

High-income students not only use rankings in their college choice, but they benefit from the rankings themselves. To boost their own rankings, colleges and universities naturally seek students with the strongest "beginning characteristics," such as GPA and SAT scores. Not surprisingly, these student selection indicators are directly correlated to the students' socioeconomic status (Meredith 2004). These indicators play an exaggerated role in the index scores of many national and international rankings. For example, 14 ranking systems from around the world incorporate some form of beginning characteristics into their calculus. Among these, four (*Guardian University Guide*, *AsiaWeek*, Education18, and *US News*) give these scores substantial weight – at least 15% (Usher and Savino 2006). Given this prominence by the rankings, universities strive to maximize their beginning characteristics, as evidenced by the increasing use of merit scholarships to recruit incoming students much to the detriment of lower SES students. Clearly, rankings stress what is already emphasized in university admissions and greatly favor students from more privileged backgrounds.

All of these organizational behaviors (e.g., gamesmanship, mimicry, recruiting high ability students, etc.) tend to exacerbate the Matthew Effect in the competitive forces in higher education. Although wonderful news for the strongest students and the strongest institutions, the consequences for student access, choice, and opportunity tend to be particularly negative for low-income and minority students (Clark 2007).

Similar to the isomorphic effect rankings have on institutional practices, rankings are also contributing to the homogenization of the socioeconomic composition found in most universities.

10.7 Lessons for University Leaders on the Efficacy of Leverage Rankings

This chapter suggests a number of lessons relevant to university leaders considering whether to and how to attempt affecting a change in their university's ranking. First, it is apparent that rankings – however "Mickey Mouse" – are here to stay and represent social constructs with real and lasting consequences. The nebulous nature of measuring higher education quality is quite consistent with the attention rankings have received from prospective students and other external constituents. The decision to simply ignore rankings can no longer be considered conscientious and will likely have consequences on any institution. However, these consequences are likely to be much greater for universities near the top of the rankings – regardless of the lingering effect of reputation – which suggests even these organizations will gently descend the rankings ladder.

Second, given the documented value of reputation – a key component of many ranking schemes – there is substantial rationale to improve a poor university reputation or protect an existing positive reputation. Granted, a boost in rankings can provide a means to improve a reputation, and vice versa, but the reputational criteria utilized in contemporary ranking schemes poorly represent institutional quality. It seems more likely that the reputational value currently found in rankings reflects the ability to charge tuition premiums and/or pay for the right to recruit and enroll high quality students at heavily discounted tuition.

Third, a large number of institutions have responded to rankings by either incorporating gaming techniques – manipulating what they can to achieve short-term improvement – or feature aspirational rankings into their organizational strategy. The number of institutions pursuing these tactics should give pause to leaders at other universities. Not everyone can be in the top 25. The rush to join the "front page" of the rankings, even given the increased number of applicants accompanying such a feat, is likely to result in many universities falling far short of their goal, even after investing substantial resources into such a plan.

Finally, any university attempting to leverage its ranking should give due consideration to the demonstrable equity concerns associated with such approaches. Current and historical studies on the topic document again and again demonstrate that, while higher rankings may be likely to produce more and better applicants, these prospective students are rarely distributed evenly across the SES spectrum. Instead, higher rankings are usually strongly correlated with less access for students from historically underrepresented populations. If the university attempting a higher ranking is public, or pursues a mission that is inclusive of these students, substantial thoughtfulness of this latent consequence is a prerequisite.

We suggest the following advice for university leaders considering the efficacy of raising their institution's ranking:

Recognize the inevitability of rankings and the constraints they impose on universities. Given the ubiquity of rankings and the attention paid to them by external and internal constituents, a "head in the sand" approach will surely fail. That said, do your homework and completely understand the variables being used in the rankings that have consequences for your university. Which variables provide some room for opportunity for your institution? It is likely that there will be some "low hanging fruit" that can be harvested from the rankings, but unless such a harvest will produce significant movement – from one tier to another, perhaps – don't expect long-term results. Identify what kind of movement is possible and consequential, given the university's mission and resources.

Avoid the allure of rankings. (see Teichler, this issue, for more details). It is common for university leaders to define their strategic plans and vision statements with ranking objectives as well as make aspirational statements related to rankings. University leaders, however, should recognize that rankings are not dynamic indicators. Rather, they more reasonably signal the rigid stability of the status quo in higher education. There is ample evidence that very few universities have moved up in the rankings and sustained this newfound position. The empirical evidence on the subject indicates that, while movement may be possible and even important if it affects perceptions of reputational quality, the quest for a higher ranking is much more likely to result in something less than success.

Recognize the importance of and buttress the university's reputation. Rankings tend to measure similar things: faculty resources, student quality, research outputs, etc. Reputations in higher education can be built upon broader variables, such as connections to the community, roles in local and regional economic development, and a commitment to mission (even if that mission is not valued by rankings indicators). There are many universities that enjoy strong reputations, with internal and external constituents, as a result of leveraging a specific niche. Although the path is not prescribed in common ranking guides, if a higher ranking is out of your university's reach, recognize that building a better reputation is valuable and entirely possible.

Beware the isomorphic grip of globalization. The criteria in the early ranking systems of the 1980s and 1990s instigated a new struggle between colleges and universities for students, faculty, facilities, and endowments. Although this competition arguably creates winners and educational improvements as well as losers and gross inefficiencies, it definitely carries significant consequences for those who participate. The more recent addition of international ranking systems will only intensify this arms race between institutions and further divide the haves and have-nots, especially as globalization increases its reach to all corners of the academic world. As institutions enter global competition for resources, they find themselves at the mercy of a cutthroat winner-takes-all campaign and the resulting inequalities can have devastating effects on academic institutions and their constituencies.

References

Barney, J. (1991). Firm resources and sustained competitive advantage. *Journal of Management, 17*(1), 99–120.

Bastedo, M., & Bowman, N. (2009, April). *College Rankings as an Interorganizational Dependency: Do resource providers respond to U.S. News & World Report*? American Educational Research annual meeting, San Diego.

Benedictine University. (2009, October 2). Retrieved from http://www.benedictine.edu

Bowman, N., & Bastedo, M. (2009). Getting on the front page: Organizational reputation, status signals, and the impact of *U.S. News and World Report* on student decisions. *Research in Higher Education, 50*(5), 415–436.

California Institute of Technology. (2009). *Caltech high in world rankings*. Retrieved from http://pr.caltech.edu/periodicals/336/articles/Volume%204/11-18-04/rankings.html

Chatterji, A. K., & Toffel, M. W. (May 2008). Shamed and able: How firms respond to being rated. *Harvard Business School* working paper 08–025, 1–45

Clark, M. (2007). The impact of higher education rankings on student access, choice, and opportunity. In Institute for Higher Education Policy (Ed.), *College and university ranking systems: Global perspectives and American challenges* (pp. 35–47). Retrieved from http://www.ihep.org

Deephouse, D. L. (2000). Media reputation as a strategic resource: An integration of mass communication and resource-based theories. *Journal of Management, 26*, 1091–1112.

Denver Post. (2009, October 2). *Metro: Great teachers or your money back*. Editorial. Retrieved from http://www.denverpost.com/opinion/ci_13466079

Devinney, T., Dowlings, G. R., & Perm-Ajchariyawong, N. (2008). The *Financial Times* business school rankings: What quality is this signal of quality? *European Management Review, 5*(4), 195–208.

Dichev, I. (2001). News or noise? Estimating the noise in the *U.S. News* university rankings. *Research in Higher Education, 42*(3), 237–266.

DiMaggio, P. J., & Powell, W. W. (1991). The iron cage revisited: Institutional isomorphism and collective rationality. In P. J. DiMaggio & W. W. Powell (Eds.), *The new institutionalism in organizational analysis*. Chicago: University of Chicago Press.

Doane College. (2009). *4 year guarantee*. Retrieved from http://www.doane.edu/Academics/Programs/4Year_Guarantee/

Ehrenberg, R. G. (2002). Reaching for the brass ring: The *U.S. News & World Report* rankings and competition. *The Review of Higher Education, 26*(2), 145–162.

Ehrenberg, R. G., & Smith, C. L. (2003). The sources and uses of annual giving at selective private research universities and liberal arts colleges. *Economics of Education Review, 22*, 223–235.

Elsbach, K. D., & Kramer, R. M. (1996). Members' responses to organizational identity threats: Encountering and countering the business week rankings. *Administrative Science Quarterly, 41*(3), 442–476.

Espeland, W. N., & Sauder, M. (2007). Rankings and reactivity: How public measures recreate social worlds. *The American Journal of Sociology, 113*(1), 1–40.

Gioia, D. A., & Thomas, J. B. (1996). Identity, image, and issue interpretation: Sensemaking during strategic change in academia. *Administrative Science Quarterly, 41*(3), 370–403.

Golden, D. (2007, March 2). Math lessons: to boost donor numbers, colleges adopt new tricks – sinking alumni stats, zeal for rankings spur rate inflation. *Wall Street Journal,* A1.

Graham, H. D., & Diamond, N. (1997). *The rise of American research universities*. Baltimore: Johns Hopkins University Press.

Hall, R. (1992). The strategic analysis of intangible resources. *Strategic Management Journal, 13*(2), 135–144.

Hazelkorn, E. (2007). The impact of league tables and ranking systems on higher education decision making. *Higher Education Management and Policy, 19*(2), 1–24.

Hazelkorn, E. (2009). Rankings and the battle for world-class excellence: Institutional strategies and policy choices. *Higher Education Management and Policy, 21*(1), 1–22.

Hossler, D. R., & Gallagher, K. S. (1987). Studying student college choice. A three-phase model and the implications for policy-makers. *College and University, 62*(3), 207–221.
Hossler, D., & Litten, L. H. (1993). *Mapping the higher education landscape*. New York: College Entrance Examination Board.
Hossler, D., Schmit, J., & Vesper, N. (1999). *Going to college: How social, economic, and educational factors influence the decisions students make*. Baltimore: Johns Hopkins University Press.
Jaschik, S. (2008, January 15). *A college cuts tuition – and ends haggling.* Insidehighered.com. Retrieved from http://www.insidehighered.com/news/2008/01/15/blackburn
Jaschik, S. (2009, September 14). *A calculation that doesn't add up.* Insidehighered.com. Retrieved from http://m.insidehighered.com/news/2009/09/14/usnews
Klein, B., Crawford, R. G., & Alchian, A. (1978). Vertical integration, appropriable rents, and the competitive contracting process. *Journal of Law and Economics, 21*, 297–326.
Labi, A. (2010, June 27). Germany pursues excellence over egalitarianism. *Chronicle of Higher Education.* Obtained online at http://chronicle.com/article/Germany-Pursues Excellence/66048/
Leslie, L. L., & Ramey, G. (1988). Donor behavior and voluntary support for higher education institutions. *Journal of Higher Education, 59*, 115–132.
Loyola University of New Orleans Loyola University 2009 Strategic Plan (2009). Retrieved from http://www.loyno.edu/presidentsoffice/documents/loyola-strategic-plan-approved-20090515.pdf
Martins, L. L. (2005). A model of the effects of reputational rankings on organizational change. *Organizational Science, 16*(6), 701–720.
McCormick, R. E., & Tinsley, M. (1990). Athletics and academics: A model of university contributions. In B. L. Goff & R. D. Tollisson (Eds.), *Sportometrics*. College Station: Texas A&M University Press.
McDonough, P. M. (1997). *Choosing colleges: How social class and schools structure opportunity*. Albany: State University of New York Press.
McDonough, P. M., & Calderone, S. (2006). The meaning of money: Perceptual differences between college counselors and low-income families about college costs and financial aid. *The American Behavioral Scientist, 49*, 1703–1718.
McDonough, P. M., Antonio, A. L., Walpole, M. B., & Pérez, L. X. (1998). College rankings: Democratized college knowledge for whom? *Research in Higher Education, 39*(5), 513–537.
Meredith, M. (2004). Why do universities compete in the ratings game? An empirical analysis of the effects of the *U.S. News and World Report* college. *Research in Higher Education, 45*(5), 443–461.
Meyer, J. W., & Rowan, B. (1977). Institutionalized organizations as myth and ceremony. *The American Journal of Sociology, 83*(2), 340–363.
Monks, J., & Ehrenberg, R. G. (1999). The impact of U.S. News & World Report college rankings on admissions outcomes and pricing policies at selective private institutions. *National Bureau of Economics Research,* working paper 7227.
Parloff, R. (1998). Who's number one? And who's number 52, 91, and 137. *American Lawyer,* p. 5.
Pryor, J. H., Hurtado, S., DeAngelo, L., Palucki, L., Tran, B., & Tran, S. (2008). *The American freshman: National norms for fall 2009*. Berkeley: University of California Press.
Rindova, V. P., Williamson, I. O., Petkova, A. P., & Sever, J. M. (2005). Being good or being known: An empirical examination of the dimensions, antecedents, and consequences of organizational reputation. *Academy of Management Journal, 48*(6), 1033–1049.
Roberts, P. W., & Dowling, G. R. (2002). Corporate reputation and sustained superior financial performance. *Strategic Management Journal, 23*(12), 1077–1093.
Sauder, M., & Espeland, W. N. (2009). The discipline of rankings: Tight coupling and organizational change. *American Sociological Review, 74*(1), 63–82.
Sauder, M., & Lancaster, R. (2006). Do rankings matter? The effects of *U.S. News & World Report* rankings on the admissions process of law schools. *Law & Society Review, 40*(1), 105–133.
St John, E. P., Asker, E., & Hu, S. (2001). The role of finances in student choice: A review of theory and research. In M. B. Paulsen & J. C. Smart (Eds.), *The finance of higher education: Theory, research, policy, and practice*. New York: Agathon.

Steinberg, M. P., Piraino, P., & Haveman, R. (2009). Access to higher education: Exploring the variation in Pell Grant prevalence among U.S. colleges and universities. *The Review of Higher Education., 32*(2), 235–270.

Texas Tech University Strategic Plan. (2005). Retrieved from http://www.ttu.edu/stratplan/

U.S. News & World Report. (2010). *Best colleges* 2010. Retrieved from http://colleges.usnews.rankingsandreviews.com/best-colleges

University of California, Berkeley. (2007). *College of Engineering. Rankings archive.* Retrieved from http://www.coe.berkeley.edu/about/rankings-archive.html#ugrad09

University of Kwaxulu-natal, Strategic Plan. (2007). Retrieved from http:www.ukzn.ac.za/about-us/strategicplan.pdf

University of Melbourne. (2010). *About the University of Melbourne*. Retrieved from http://www.futurestudents.unimelb.edu.au/about

Usher, A., & Savino, M. (2006). *A world of difference: A global survey of university league tables*. Ontario: Educational Policy Institute, pp. 1–38.

Weber, M. (1958). *The protestant ethic and the spirit of capitalism iron cage*. New York: Charles Scribner's Sons.

Wedlin, L. (2007). The role of rankings in codifying a business school template: Classifications, diffusion and mediated isomorphism in organizational fields. *European Management Review, 4*, 24–39.

Weggel, A. (2007). 5 Years tuition for the price of 4, Juniata college offers. *Chronicle of Higher Education, 54*(11), A28.

Weisbrod, B. A., Ballou, J. P., & Asch, E. D. (2008). *Mission and money: Understanding the university*. Cambridge: Cambridge University.

Chapter 11
The Institutionalization of Rankings: Managing Status Anxiety in an Increasingly Marketized Environment

William Locke

11.1 Introduction

This chapter is part of a larger effort to understand how higher education institutions (HEIs) as organizations are responding to marketization, and how this influences intra-institutional relations, organizational cultures, and management styles (Locke and Botas 2009; Locke 2010). A recent study for the Higher Education Funding Council for England (HEFCE) by the author and colleagues concluded that institutional rankings were being used for a broader range of purposes than originally intended, and bestowed with more meaning than the data alone may bear (Locke 2008). The study found, in particular, that higher education institutions in England were strongly influenced by rankings in both their strategic decision-making and more routine management processes. Case study institutions reported increasing reference to the rankings by prospective students and their families and by academics seeking job opportunities. Other studies have highlighted their use by employers in the marketing of graduate jobs and the selection of candidates (Morley and Aynsley 2007). Yet, analysis of three UK national tables and two world rankings confirmed they largely reflected institutional reputation and resources rather than the quality or performance of institutions.

A higher education institution's reputation is based on how attractive it is, and therefore how selective it can be, with regard to students, academic and other professional staff, research funders, and partnerships. As higher education becomes increasingly subject to marketization, reputation becomes critical because it is regarded by universities, employers, government, and the best qualified and most mobile students as ultimately more important than quality. However, the diversion of resources toward activities that enhance institutional reputation may actually detract from the quality and performance of educational activities that are likely to be of most interest to potential students and their families. Expenditure on extensive

W. Locke (✉)
Centre for Higher Education Research and Information,
The Open University, London, UK
e-mail: W.Locke@hefce.ac.uk

J.C. Shin et al. (eds.), *University Rankings*, The Changing Academy – The Changing Academic Profession in International Comparative Perspective 3,
DOI 10.1007/978-94-007-1116-7_11,

marketing campaigns, impressive new buildings and facilities, and attracting international research "stars" are thought to be a signal of "high quality" and are, therefore, likely to increase shares in the markets for students, consultancy services, and research funds. But this may mean that money is not spent on supporting students' learning, improving educational resources, and the professional development of younger academic staff.

The interaction between rankings and marketization helps to explain why compilers and publishers have been surprised by the influence of their rankings; market mechanisms and responses to these have transformed their (not entirely innocent) attempts to provide simple and "user friendly" guides to the higher education landscape for prospective students and their families into vehicles for auditing and producing changes in performance. It also explains why attempts by higher education institutions themselves to boycott rankings have largely failed; rankings are linked with larger and more far-reaching changes in economies and society that cannot simply be rejected, and they appear to have to be, at least in part, accommodated, even where they are resisted in principle.

However, this accommodation – a form of internalization and, ultimately, institutionalization of ranking systems' rationales and processes – may produce unintended and undesirable consequences for higher education institutions. The concepts of internalization and institutionalization help to explain the processes by which ranking systems logic becomes embedded in organizational structures and procedures and established as the norm, despite initial skepticism and resistance.

This chapter first aims to situate the approach adopted here in the literature on the influence of rankings on higher education and calls for more empirical research on the ways in which individuals and entities within institutions address, understand, and handle rankings. In particular, it draws on three studies of institutions' perspectives on rankings. It then asks how we can conceptualize the ways in which these organizational members make sense of, and then respond to, rankings. Here, the approaches adopted by some US researchers seem most likely to be fruitful. The empirical core of the chapter is a re-analysis of one of the studies of institutions responding to rankings (which the author was involved in) drawing on the conceptual framework already outlined. The key findings of this re-analysis are discussed and the chapter concludes by calling for further differentiated and conceptually-informed empirical investigation of the influence of rankings on, and within, higher education institutions.

11.2 Levels of Analysis of the Influence of Rankings on Higher Education

To date, the influence of rankings on higher education has largely been studied at the level of global policy (Marginson and van der Wende 2006, 2007; Marginson 2007, 2009) and global regulation, and in the context of national higher education systems (Wedlin 2006; Locke et al. 2008; Locke 2008; IHEP 2009; Teichler in Chap. 4).

Reference is often made to national governments' desire to foster "world class" research-intensive universities as a source of comparative economic and status advantage (King 2009). An "emerging global model" of the research university has been developed to characterize the kind of institution that appears at the top of the world rankings (Mohrman et al. 2008). The German Exzellenzinitiative (2005) is a key example identified in the literature of the influence of the global rankings on the thinking of national governments. The shift from egalitarianism to competition this represents (IHEP 2009) is leading, it is argued, to the concentration of resources and reputation, the undermining of meritocracy, and the increasing vertical stratification of higher education systems (Teichler 2008). Rankings themselves are becoming instruments of national and institutional policy (Salmi and Saroyan 2007; Hazelkorn 2009). In Japan, for example, higher ranked universities receive more attention from the central government (as distinct from the Ministry), including the allocation of funds (IHEP 2009). There is also evidence from the USA of rankings influencing research and development funds from government (Bastedo and Bowman 2011). Their use as indicators of performance lends themselves to institutional benchmarking and determining internal resource allocation.

The few investigations of the influence of rankings on individual higher education institutions have tended to focus at the level of the whole institution (Locke et al. 2008; Cyrenne and Grant 2009), often from the perspectives of institutional leaders (Hazelkorn 2007, 2008, 2009; Morphew, in Chap. 10), or specific disciplines and specialist schools (Elsbach and Kramer 1996; Walpole 1998; Martins 2005; Sauder and Lancaster 2006; Wedlin 2006; Espeland and Sauder 2007; Sauder and Espeland 2009). They tend not to distinguish between types of institution (except, perhaps, by broad ranking position) and the different levels of resources these can draw on to resist external pressures (or exploit competitive opportunities) such as rankings, and do not chart over time the evolving effects of these influences. It is not entirely a coincidence that rankings also largely focus at the level of the whole institution rather than the discipline or department, tend not to distinguish between different types of institution (except in broad terms) and the resources these can draw on, and do not track institutions' performance over time, preferring a "snapshot" approach to evaluation.

Perhaps we should not be surprised that these initial analyses of the influence of rankings mirror the main features of ranking systems themselves. However, it is the differential responses to rankings of distinct types of institution and the different parts within an institution, the relations between internal units and university-wide management, and the activities within institutions – together with the unfolding of these responses over time – that have yet to be explored empirically to any great extent and in any detail. One small example may help to illustrate why this level of analysis is significant and where someone is situated in a higher education institution will tend to influence the nature and degree of their response to rankings. The further away from the "frontline" – in the sense of being at the interface between the institution and its various markets (for example, student recruitment, fundraising, marketing) – the more equivocal tends to be an individual's response to rankings.

11.3 Rankings and Marketization

One impediment to assessing the influence of rankings on the internal operations of higher education institutions is the difficulty of separating their influence from the general effects of competitive forces and other factors on higher education (IHEP 2009), such as improved organizational management, quality enhancement, and the greater value placed on research in science than in other disciplines. In particular, rankings feed on governments' efforts to improve public accountability and institutional performance by means of published indicators of teaching and research quality, student satisfaction, and graduate employment, for example.

> Rankings are used to underpin government exhortations to be more competitive and responsive to the marketplace and customers, define a distinctive mission, be more efficient or productive, and become world-class. (Hazelkorn 2009: 68)
>
> In Australia and Denmark, rankings are linked with institutional contracts or compacts and quality assurance, becoming a "quasi-funding instrument" (Hazelkorn 2009: 70).

However, rankings also produce perverse changes in institutions' academic priorities that may simultaneously contradict government and institutional policies, for example, on widening access to disadvantaged students who have not had the opportunity to achieve high qualifications, promoting greater diversity among institutions' missions, and maximizing the socio-economic impacts of research. Some of those institutions ranked lower than expected are, for example, having to develop survival strategies to bolster demand in key overseas markets damaged by league tables, revive their internal morale and public confidence, and spend more on marketing and publicity to restore their image (Hazelkorn 2008). Many of these actions can be traced directly to the influence of rankings (Locke et al. 2008). But they may often be short-term oriented, and even potentially self-damaging actions (Gioia and Corley 2002; Martins 2005).

Rankings exert pressure to "conform and perform" to their criteria (Gioia and Corley 2002) because institutions perceive their key stakeholders are, themselves, influenced by them. Despite the relative paucity of evidence of this, institutions assume that high achieving undergraduate and graduate applicants, graduate employers, talented researchers, research funders, potential partners, foreign scholarship awarding bodies, government agencies, and donors are swayed by the rankings. They anticipate the impact this might have on their access to resources (Martins 2005). This reaction to rankings illustrates the preparedness of universities to compete (IHEP 2009) and the extension and embedding of market logic in higher education institutions (Sauder and Espeland 2009). Rankings are also being used as marketing tools (Grewal 2008) and for image management (Gioia and Corley 2002).

In this way, rankings serve to reinforce the effects of broader, market-based, and competitive forces in higher education (Clarke 2007) and are ratcheting up the level of competition between institutions. In federal systems, they even accelerate competition between states or länder (IHEP 2009). Ranking systems are helping to transform higher education institutions into strategic corporations, engaged in

positional competition to close the gap between their current and preferred rank (Hazelkorn 2009). They encourage institutions to become more business-like and respond to their "customers" promptly. It is no coincidence that rankings have proliferated (and been least contested) in the most marketized higher education environments, such as the US, Japan, the UK, and Australia. The influence of market behavior explains the failure of collective resistance to rankings by higher education institutions in these territories.

11.4 Institutional Perspectives on Rankings

There are few empirical studies of higher education institutions' views on rankings and their influence. Three recent studies are reviewed in this section.

The international survey of higher education leaders and senior managers, on the impact of rankings on strategic and operational decision-making and choice supported by the OECD, undertaken by Hazelkorn (Hazelkorn 2007, 2008, 2009) is the most extensive study.

In outline, Hazelkorn's key findings were:

- Fifty-seven percent of respondents thought the impact of league tables and ranking systems had been broadly positive on their institution's reputation and broadly helped its development, in attracting students, forming academic partnerships, collaboration, program development, and staff morale.
- The methodologies, however, were described as simplistic and lacking transparency by many respondents, although the nature of the responses depended to some extent on the institution's position in the tables, with those among the lower rankings criticizing league tables for creating "a cycle of disadvantage."
- Fifty-six percent had a formal internal process for reviewing their rankings.
- The majority (63%) had taken strategic or academic decisions or actions, incorporating the outcomes into their strategic planning mechanisms, using them to identify weaknesses and resolve institutional problems, develop better management information systems, professionalize services such as admissions, marketing and careers, and allocate resources accordingly.
- A significant minority (40%) used rankings to inform their decisions about entering into partnerships with other institutions and joint academic and exchange programs.
- Seventy-six percent monitored the performance of other higher education institutions in their country, including considering this before starting discussions about collaboration. However, most institutional users tended to draw broad brushstroke conclusions from them, using the results to reassure themselves.

The study by the author and colleagues (Locke et al. 2008) included a survey of English higher education institutions that was informed by Hazelkorn's initial findings (Hazelkorn 2007). What follows is a summary of the major findings of this survey undertaken in 2007.

11.4.1 Perceptions of Rankings

Respondents agreed that league tables often reflect "idiosyncratic views" of what constitutes "a good university" that are often at considerable variance from institutional and governmental policies. There was also a relatively high agreement that rankings may affect institutional reputations and even damage these beyond repair. Despite some respondents stating that they had increased their communication with the league table publishers, it was widely felt that institutions did not have enough influence on the compilers of rankings and the methodologies used. League table publishers were thought to benefit the most from the production of league tables, followed by students, the Government, employers, higher education institutions (the latter three were relatively closely ranked together), and, finally, academic staff.

When asked whether their institution was ranked broadly where they would expect it to be, respondents were almost equally divided, with 44% answering "Yes" to this question and 52% saying "No." The respondents, who stated that their institution was ranked differently to where they would expect it to be, were asked to indicate where they thought they should be ranked. Almost half of these respondents indicated that they thought they should be ranked at least ten places higher. Just over a quarter stated that they should be ranked at least 20 places higher, and a fifth at least 30 places higher.

11.4.2 The Importance of Achieving Good Rankings

There was a high level of agreement that achieving good rankings was important. The areas perceived to be influenced the most by league tables were the general reputation of the institution, followed by student recruitment, employers and the building of networks and alliances, then key performance indicators (KPIs), and finally the recruitment of academics and attracting financial resources (ranked similarly).

11.4.3 League Tables as a Driver for Change

Most survey participants stated that they had responded to league tables in some way. However, institutions stressed that league tables were not driving institutional agendas and that "actions are not developed as a knee-jerk to tables," as one respondent put it.

One of the most frequently mentioned responses was to undertake analysis of institutional positioning in league tables and the methodologies used. Close to half of those institutions who stated that they had responded to league tables, indicated that this analysis was being carried out in their institution. A high number indicated that they had established dedicated working parties to consider

league tables and the issues associated with them. About a third of respondents indicated that league table publications had made their institution consider how data returns were made and prompted them to establish mechanisms to assure the quality of information provided to the national statistics agency.

The National Student Survey (NSS)[1] was frequently mentioned as one of the indicators used in league tables which had prompted the institution to take action. A number of respondents stated that initiatives to analyze or address NSS results had been taken. Some respondents indicated that they had enhanced communication with league table compilers to try to influence methodologies, take advice on how to optimize data returns or just better understand how rankings are compiled. A majority of institutions also stated they used the league table variables to review key performance indicators. Other areas, in which institutions had responded, included marketing and communication, and internal discussions concerning league tables.

Institutions were found to have made most changes due to league tables in promotion and marketing activities, and in the way they made decisions about how to submit data, media relations, and key performance indicators or targets. The areas which had been influenced the least were arguably core institutional areas such as staff recruitment policies, course offering and content, and research. Respondents were asked to give examples of changes in strategies, policies, and priorities (including resource-allocation decisions) which could be directly attributed to the institution's response to the published rankings. More than half of respondents either did not answer the question or stated that they had not made any changes as a direct result of rankings; some of them emphasized that changes to strategies or policies were not driven by league tables.

Amongst institutions which stated they had implemented changes as a result of rankings, a less clear pattern was detected compared with answers on how they had responded to league table publication. The way data are returned to national agencies was again identified as one of the areas receiving the most attention. The NSS results were highlighted as an important influence, with a few institutions stating that they had made changes to academic facilities (for example, library and IT infrastructure) and student support services. Other changes identified by a small number of institutions included improved careers services, and initiatives to address the student: staff ratio, the proportion of good degrees awarded, research, the quality of teaching and learning, and public relations and marketing.

Institutions were asked whether the changes implemented would have happened regardless of the league tables and, if so, whether the rankings gave impetus to implementing the changes. Most respondents either agreed or strongly agreed that

[1] The National Student Survey (NSS) of final year undergraduate students aims to inform prospective students and their advisers in choosing what and where to study. The questionnaire consists of six "areas" covering: Teaching, Assessment and feedback, Academic support, Organization and management, Learning resources, and Personal development. There is also an "overall satisfaction" question about how satisfied the respondent is with the quality of their course. The first full-scale NSS took place in England in 2005.

this was the case. Only a minority agreed that rankings created pressure to take actions that they would rather not have taken.

A more recent study by the US Institute for Higher Education Policy (also involving Hazelkorn) was based on interviews and focus groups of senior administrators, academics, and students at 20 higher education institutions of various kinds in four countries. The report of this study (IHEP 2009) helpfully categorizes the influence of rankings on institutional decision making into the following five primary but interrelated areas:

Strategic positioning and planning: An institution explicitly setting a goal of improving the rankings, using resource allocation, recruitment, and other decision making to pursue this goal. Examples given from Germany included formulating strategic ambitions in terms of ranking positions, benchmarking with peer institutions, using ranking position to lobby Länder for additional resources, and informing the selection of rectors.

Staffing and organization: Reconfiguring an institution's organizational structure, merging smaller units, creating specialist units to address particular aspects relevant to rankings, and recruiting and promoting staff on the basis of anticipated improvement in the rankings. Specific examples included establishing groups to review rankings and manage related information and data, setting up international student recruitment offices, modernizing admissions and marketing functions, and increasing the salaries of permanent faculty while restricting the number of these positions.

In Germany, the authors reported that:

> Competition will likely lead to changes in the academic profession and work practices; new performance measures and recruitment practices will challenge the traditional power of the professoriate and individual professors (IHEP 2009: 17).

Quality assurance: The review and revision of the curriculum, academic programs, teaching, learning and assessment practices, the use of communications and information technology, and graduation rates, for example. In those countries where quality assurance or accreditation processes are relatively underdeveloped, ranking systems can provide a convenient – if unsophisticated – substitute. Government or state funding decisions may be informed by these.

Resource allocation and funding: The distribution of resources within institutions may be partly based on rankings, including administrators' and faculty pay, investment in high profile projects such as buildings, and expenditure on students. Strategies adopted for fundraising, for example from alumni, may be influenced by rising or falling ranking positions.

Admissions and financial aid: This is the area where there is most evidence of the influence of ranking systems on higher education institutions' actions, especially in the USA. Examples include: efforts to lower the acceptance rate; biases arising from the measures of selectivity used by compilers, such as SAT scores; and the greater use of merit aid to attract highly qualified students leading to reductions in the availability of need-based aid.

In several of the countries included in this study, respondents reported a high ranking consciousness throughout institutions, a sense of foreboding and concern about the influence of rankings and their impact on key stakeholders. Among Australian academics, for example:

> The overall effect of rankings is to generate anxiety, resulting in a great sense of pride if the university is doing well but anger or shame if it is not (IHEP 2009: 12).

11.5 Making Sense of Rankings in Institutions

These policy-inspired studies provide useful empirical evidence at the institutional level of analysis. But, how can we conceptualize the ways in which higher education institutions and their members internalize the logic of ranking systems, and their influence becomes institutionalized in organizational processes and structures? A number of US researchers have begun to tackle this, although largely focusing on how rankings influence specialist academic organizations, such as law schools and business schools.

Espeland and Sauder have employed the concept of "reactivity" – or how people change their behavior in reaction to being evaluated, observed, or measured:

> The proliferation of quantitative measures of performance is a significant social trend that is fundamental to accountability and governance…(which) can initiate sweeping changes in status systems, work relations, and the reproduction of inequality. (Espeland and Sauder 2007: 2)

A reactive measure is one that changes the phenomenon it is designed to evaluate, because those who are being evaluated begin to concentrate on the measure rather than the activity. Espeland and Sauder conceptualize the nature of reactivity as patterns that shape how people within organizations make sense of things and how they interact with rankings, each other, and other institutions. They identify two of the most important mechanisms that induce reactivity, "the self-fulfilling prophecy," and "commensuration."

Self-fulfilling prophecies are:

> Processes by which reactions to social measures confirm the expectations or predictions that are embedded in measures or which increase the validity of the measures by encouraging behavior that conforms to it. (Espeland and Sauder 2007: 11)

For example, rankings exaggerate minor differences between higher education institutions, creating artificial distinctions that may become real. Several US researchers have found that ranking position can influence the number and quality of applications a school receives and its yield (the proportion of accepted students who attend a school) (Monks and Ehrenberg 1999; Meredith 2004; Sauder and Lancaster 2006; Bowman and Bastedo 2009). This, in turn, will change the measure of selectivity used in the *US News and World Report* (USN) ranking. More self-conscious forms of self-fulfilling prophecies include adopting improved ranking positions as an explicit institutional goal and using rankings to characterize and market the institution to external audiences.

Commensuration is characterized by:

> The transformation of qualities into quantities that share a metric…(It) shapes what we pay attention to, which things are connected to other things, and how we express sameness and difference. (Espeland and Sauder 2007: 16)

Ranking systems, for example, simplify complex information, decontextualize it so that it can be organized and integrated in particular ways, and eliminate huge amounts of other, qualitative, information that cannot be assimilated within the system. Many of the factors most important to prospective students are undervalued or entirely excluded by compilers. But, because numbers are depersonalized, they appear more authoritative and definitive and, once they are decontextualized, they can be put to new purposes in new contexts, such as the internal management of higher education institutions.

Espeland and Sauder also identify three important effects of reactivity: the redistribution of resources, the redefinition of work, and the proliferation of gaming strategies. They describe, for example, how law schools offer merit scholarships to students with high test scores in order to improve their selectivity results in the USN ranking, and send expensive marketing brochures to peers and practitioners in an attempt to improve their chances in the same ranking's reputation survey. Careers services personnel expend more effort tracking down their graduates' employment status at the expense of time spent on counseling students, running employment seminars, and interacting with graduate employers. Admission staffs are under pressure to base decisions on rankings considerations rather than professional judgments. Finally, law graduates in non-legal jobs are recorded as "employed" to include them in the statistics, and students with low SAT scores are classified as "part-time" or "probationary" to exclude them.

The authors argue that these two means of inducing reactivity tend to produce effects at different rates: commensuration can have an immediate effect because it changes the form of information; and self-fulfilling processes, however, may only emerge gradually, as it takes time for people to alter their expectations and modify their behavior accordingly:

> Both mechanisms of reactivity will produce varied changes over time; initial responses will be different from those coming later, after rankings have diffused more broadly or become naturalized. (Espeland and Sauder 2007: 23)

Higher education institutions' initial reactions may be dismissive, but when it becomes clear that others – prospective students, their parents, graduate employers, alumni, other higher education institutions, lay governors, and government officials – are taking notice of rankings, managers start to treat them more seriously. They may then seek to understand the ranking systems and how their institutions' data are employed in the calculation of ranking positions. They may criticize the criteria used by compilers, seek to obtain modifications and attempt to "adjust" the information they present. Later on, institutions might start to invest in improving their rank positions, adjusting decision making to take account of the effects on rankings, using them for promotional purposes, and incorporating them in strategic planning.

In detailing how the most important mechanisms of inducing institutions' reactivity to rankings, Espeland and Sauder have started to explore the means by which organizational members begin to internalize the logic of rankings and how their influence becomes institutionalized in processes, systems, and structures over time. This is not to suggest that these are smooth, uncontested, or inevitable changes. It may even be that, while recognizing and criticizing the transforming influence of rankings, institutional members feel compelled – by reduced public funding, market forces, institutional leaders, or government or state policy, for example – to facilitate and extend their effects (Gioia and Corley 2002). Low rankings, in particular, can lead to a – seemingly unstoppable – downward spiral of negative impacts on funding, student enrollment, staff recruitment, and research capability (Walpole 1998).

Other US researchers have investigated how rankings can be a threat to organizational members' perceptions of their institution's identity and their beliefs about its standing relative to other institutions. Aspects of an organization's identity that are not included in rankings become perceived as less important and even irrelevant as indicators of the institution's performance or quality (Elsbach and Kramer 1996). Staff members and others associated with the institution (such as students, alumni, and governors) may experience a kind of "identity dissonance" between their perception of the organization and the picture that is presented by the publishers of rankings, even in highly-ranked institutions. In order to resolve this dissonance, members may make excuses or provide justifications. But they may also attempt to attenuate or mitigate identity threats by emphasizing other ways in which the organization is intrinsically good or functioning well, or by highlighting their institution's membership of alternative comparison groups (for example, higher education institutions in the same region, those established at around the same time, or those of a similar type).

11.6 Institutions Responding to Rankings

Espeland and Sauder have gone on to develop a more sophisticated understanding of how organizations respond to rankings in a later article (Sauder and Espeland 2009). They have developed a Foucauldian analysis to explore how rankings have permeated US law schools so extensively, preventing them from resisting their influence, despite vociferous protests. Drawing on Foucault's conception of discipline, they argue that it is difficult for institutions to buffer these institutional pressures because of the ways in which organizational actors tend to internalize external pressures and become self-disciplining. This internalization is fostered by the anxiety produced by rankings, the resistance they provoke, and the attraction for administrators and others of trying to manipulate them.

> Analyzing rankings as a form of disciplinary power reveals that rankings, through processes of surveillance and normalization, change how internal and external constituencies think about the field of legal education. These new understandings of legal education, in turn, encourage schools to self-impose the discipline that rankings foster. Rankings also offer

> external audiences a means for compelling law schools to meet their demands. Rankings change perceptions of legal education through the incentives that are simultaneously seductive and coercive.
>
> …..
>
> Foucault's focus on internalization explains the appeal that underlies these measures for members and outsiders alike. Specifically, Foucault shows how coercive disciplinary pressures devolve into forms of "self-management" … that amplify institutional influences by changing members' perceptions, expectations, and behavior. In addition, this approach helps explain how efforts to control rankings, whether through strategic manipulation or resistance, propel the institutionalization of rankings and extend their power. (Sauder and Espeland 2009: 64)

These complex processes of accommodation involve "an assortment of actors who struggle to reconcile their sense of themselves as professional educators with an imposed market-based logic of accountability" (Sauder and Espeland 2009: 66). Their reactions may vary and change over time – some may try to resist and others may focus on attempts to manipulate the rankings – but, the authors argue, rankings become naturalized and legitimized as arbiters of status for the vast majority of institutions and their members.

Sauder and Espeland's conceptual framework allows them to analyze the influence of rankings in a dynamic and nuanced way. It highlights how rankings are not simply imposed on institutions of higher education from outside, and that resistance and manipulation are possible. As in their earlier article, their approach also acknowledges how institutions' responses evolve over time, and how rankings seduce as well as coerce. This is an important antidote to those analyses that underestimate the power of institutions to respond actively to environmental forces and that assume they react passively to external pressures. Ultimately, though, even resistance and manipulation lead to the insinuation and normalization of ranking systems logic, and there is little sense in Sauder and Espeland's analysis of any positive or constructive effects for some institutions and for some stakeholders. Moreover, the foucauldian concepts they employ do not lend themselves to exploring the reverberations of rankings within institutions, for example, how they are used by governing bodies and senior management to drive change, by particular disciplines to argue for more resources and by individual academics to enhance their career prospects. Also, the language of "impact" and "buffer" is one dimensional. Effectively, it only countenances two possible responses to the normative pressure of rankings: conformity or resistance – and ultimately, anyway, resistance turns out to be "futile." Yet, conformity to rankings is not an inevitable or prescribed process.

There have been few attempts in the literature so far to understand why organizations vary in their responses to rankings and why some are more likely than others to change as a result of such external pressures (Martins 2005). Analysis needs to examine not just the differences in the degree of change, but also variations in the nature of that change. Rankings also have specific effects, particularly in relation to those indicators that influence the resources and reputation that an institution can attract (for example, research funding and recognition, student recruitment, tuition

fees, student opinion, and peer review). The relevance and effects of these indicators vary, depending on the type of higher education institution. Particular measures are included in some rankings and not in others, and some indicators are not included in any commercially published rankings (for example, measures of access to non-traditional groups of students). Whether, where, when, and how rankings serve as an incentive for change may depend on the academic unit, the nature of the ranking, and the length of time during which a lower than expected (or desired) ranking is experienced (Walpole 1998). The answers to these questions will depend on careful empirical analysis and cannot simply be "read off" from the generic features of rankings and the common reactions of higher education institutions and their members. The remaining sections of this chapter attempt to make a start in this more differentiated approach to understand the influence of rankings on higher education institutions in a specific territory, England.

11.7 Analysis of English Case Studies

This section presents the results of a re-analysis of the detailed findings of the case studies of six higher education institutions in England included in the appendix to a report to the Higher Education Funding Council for England (HEFCE) in 2008 (Locke et al. 2008: http://www.hefce.ac.uk/pubs/hefce/2008/08_14/08_14e.pdf). The case studies were based on semi-structured interviews with key personnel from each of the institutions, for example, representatives from the senior management, governing body, careers services, and departments concerned with communication and marketing, and domestic and international recruitment. In addition, two focus groups at school or departmental level were held in each institution where possible. Phrases and sentences quoted below are taken from summaries of these interviews and focus group discussions. The research team also analyzed documents provided by the higher education institutions and published material, for example, on web sites. The participating institutions were selected to be as representative of the different types of institution in the sector and positions in league table rankings as possible. Where relevant, this case study material is amplified by the results of an online survey of all higher education institutions in England.

The re-analysis draws on previous attempts to interpret the impact of rankings on institutional behavior (Elsbach and Kramer 1996; Walpole 1998; Espeland and Sauder 2007; Sauder and Espeland 2009), but extending these to interpret the gradual internalization of the particular logics of rankings at different levels of the case study institutions (e.g., governing bodies, senior management, school, and department) and by different parts of each university (academic, administrative, marketing, recruitment and admissions, curriculum committees, data and planning, and so on). Re-analysis of the case study evidence revealed a number of ways in which different types of higher education institution and distinct levels and parts of institutions are affected by, react to, and use rankings in various ways.

In particular, the following six interlinked themes are highlighted in this section of the chapter:

- The use of rankings in the strategic positioning of institutions, in branding and promoting themselves, and in making decisions about strategic goals
- How rankings can redefine activities as institutional personnel focus on the indicators and measures used in rankings rather than the qualities they are designed to evaluate, privileging certain characteristics above others
- How responses to rankings evolve, for example, from initial dissonance and the invoking of alternative evaluations, to attempts to understand and explain unexpected results, to efforts to produce desired ranking outcomes, and the exploitation of ranking successes in institutional promotion activities
- The influence of ranking results in the affective domain, including the impact on staff morale in institutions (and departments) ranked in different parts of the national tables, and anxiety about what other institutions are doing to improve their ranking positions
- The use of ranking logics to lever internal change, for example, tightening reporting procedures, rendering academic units accountable, and promoting competition between departments
- Attempts to manage the influence of rankings, including negotiations with compilers and efforts to mitigate conflicts between ranking logics and the social missions of institutions

11.7.1 Strategic Positioning and Decision-Making

Rankings are now one of the main mechanisms for higher education institutions to position themselves – in markets, as members of particular lobbying groups and in formulating their institutional missions and aims, for example. A case study university near the top of national tables and rising within the international rankings in the study sought, increasingly, to position itself as a global university, claiming or aspiring to be "world class" and "a global player," an attractive brand to students and academics worldwide and operating on "an international stage." It took much more notice of the global than the national rankings. Its publicity, initially coy, soon began to proclaim its ranking successes in bold terms. The marketing strategy had not been so reserved. The university had noticed the high proportion of academic respondents to the peer survey in the *Times Higher Education*'s World University Rankings from India. As a result, it worked "quite hard" at improving brand recognition in that country, to raise awareness of the university among potential survey respondents.

This self-image as a top-ranking university had also given rise to internal scrutiny. *The Guardian's* subject tables revealed a number of academic departments that were not in the top 20, which seemed "inconceivable" to the senior management of the university. A working party was established to review the rankings and data submissions to the national statistics agency. Questions were asked about whether a

"top-10" university should be doing some of the things it was doing, such as relying on the national "clearing system" for matching unplaced applicants with unfilled places just prior to the start of courses. In one department, entry requirements were raised but also broadened as a result of a review of degree programs and the way that prospective students select subjects to study at advanced secondary level.

For those institutions just below the "elite," the national tables are, perhaps, more key to strategic positioning, because they can help to distinguish a university in the minds of the public from the mass of mid-table institutions. It is important for these universities to situate themselves within "competitive sets" or clusters of institutions immediately above and below them in the rankings. One such institution sought to emulate the institutions at the top of their set or band as these have the same "background" and "heritage." The governing body and senior management spent a lot of time asking why their "peer competitors" had done better in the national tables and what they would have to do to catch them.

Indeed, it is often the lay governors of an institution that have become most exercised about ranking positions and appear more susceptible to ambitious and unrealistic expectations about where the institution could or should be positioned. League tables simplify complex processes and are familiar from other areas of competitive activity, such as sport and business. They are a handy way for lay governors to exert pressure on a university management who may seem "complacent" or constrained by academic obduracy and belligerent trade unions. In one former polytechnic, much lower than other similar institutions in the national tables, the vice-chancellor had committed to improving the university's position but, in the view of one governor, "had made a rod for his own back" and risked his own job security. In another former polytechnic, highly positioned among its peers, the governors were unwilling to bankrupt the university just to improve rankings but were keenly aware of their importance, particularly in the international student market on which the institution relied for a significant proportion of its income.

For institutions ranked in the bottom half of national league tables – focused more on teaching vocational subjects than on research – the importance of the annual NSS results was clear from our study. Recently included in UK league tables, the NSS can be a double-edged sword for institutions with limited resources. On the one hand, the Survey focuses on final year undergraduate students' views of their study experiences – an area in which teaching-focused institutions seek to excel. On the other hand, without research income and endowment funds, some institutions have difficulty improving their staff-student ratios or developing their teaching accommodation and facilities to the extent that research-intensive universities can.

11.7.2 Redefining Activities and Altering Perceptions

For some lay governors of universities in the study, league tables have introduced "a sense of the market" and of the consumer, and introduced "market discipline." They have contributed to an increasing awareness of market pressures generally in

higher education, accelerated by the introduction (in England), in 2004, of higher tuition fees in the form of student loans to be repaid on graduation. For these governors, league tables are "a short-hand descriptor" for the market and competition overall. The benchmarking made possible by league tables had highlighted, for them, the need to target resource allocation and to be forward-looking. It had focused minds on "performance management" and "research effectiveness."

For other governors, however, the tables were counter to "more sensible" and "proper" external evaluations, such as institutional quality assessment and financial audit, which they considered more suitable for the purposes of evaluating a higher education institution's activities. Accordingly, to use rankings to "make things happen" or "as a stick against the administration" would be very unwise, because "league tables have a power beyond rationality."

A senior manager in a small university college thought that league tables had highlighted how the institution had not been "terribly business-focused." "The business is education, but we haven't measured that and improved management information." For departmental staff in a former polytechnic, league tables were part of a more formalized approach to evaluation generally. "It has taken us aback and made us realize that our tacit knowledge of ourselves and our 'ranking' in the subject community may not be objective enough."

Two areas of activity subject to redefinition were common to several of the case study institutions in our study and directly related to elements of the methodologies for compiling league tables: "the student experience" and "graduate employability." The substitution of the NSS results for the increasingly out of date grades for assessed teaching quality, awarded to academic departments by the national Quality Assurance Agency, seems to have led to the reduction of teaching and learning to the six categories and 22 questions included in the Survey. For the provice chancellor of one older university, the NSS had "helped" them "to rebalance teaching with research." It was a counterweight to the periodic Research Assessment Exercise. It put the spotlight on heads of departments with poor NSS results and introduced peer pressure from other heads to make improvements. In several of the institutions, this resulted in enhancement to student support services, the building of new study facilities, and extensive staff development activities. The impact on teaching quality, the curriculum and students' learning outcomes, however, was not so apparent.

Several institutions had also responded to the indicators of graduate employment in the national league tables that are mainly based on annual surveys of full-time graduates 6 months after graduation (First Destination Returns, FDR). Careers education and job placement services had been reorganized, university-wide "employability strategies" were implemented and academics encouraged to incorporate students' work-related skills development and engagement with graduate employers into the curriculum. The quality of the FDR as an indicator is almost universally criticized for being a short-term and snap-shot measure, which does not discriminate between disciplines, differences in graduate mobility, and graduate labor market circumstances.

11.7.3 Evolving Responses

Institutions' responses to rankings changed over time. Initially, league tables may be viewed as solely a media relations issue – a success to be celebrated or a public relations crisis to be managed. Most institutions were naïve to begin with, uncomprehending of the methods by which individual tables were compiled, and unaware of the connection between the data they supplied to the national statistics agency and the outcomes of the rankings. In most of the case study institutions, the first action was to establish a "working group" to analyze the league tables, including consulting with the more forthcoming compilers, and to review how data are gathered and submitted by the institution. Common outcomes were to include data that had previously been missed out, such as departmental expenditure on subject specific libraries and those academics on fractional contracts. Higher ranked universities also had to point out to compilers that they had included them in tables for the disciplines they did not teach, due to the mis-coding of students. Subsequently, discussions about league tables tended to move from this mid-level both upward, to the senior management and governing body, and downward, to the departments. However, in several of the lower ranked case study institutions, it was evident that departmental staff remained unclear about the connection between the data they supplied to the center, the institution's submission to the national agency and the published league tables based on these submissions.

There were differences between types of institution in their initial responses. In the research-intensive university near the top of both national and global rankings, the working group was relatively junior and dealt mainly with the administrative implications of league tables – perhaps, because the senior management were keen to emphasize that decision-making was not driven by the rankings. In lower ranked institutions, the groups tended to be led by a senior manager (a deputy or pro vice-chancellor) or the issue was referred to one of the key standing committees in the institution – perhaps because senior management accepted that their institutions could no longer afford to ignore league tables.

For both former polytechnics in the study, the league tables represented "a wake-up call." Initially, these lower-placed institutions had not understood the significance of their data returns, been "too modest" or "missed a trick." Because these institutions were skeptical about the importance of league tables for the majority of their student intake – who were neither middle-class nor mobile – they had initially underestimated their broader influence on the reputation of their institution. Non-university institutions, however, that do not feature in some of the national tables, were quick to protest against their exclusion. It appeared they wanted the visibility of at least being included in "the map" of higher education presented by publishers of league tables, even if they had little hope of featuring above the bottom quartile. At this stage, visibility was more important than parity.

The global rankings were largely ignored by those higher education institutions ranked below the top quartile of the national tables, for obvious reasons. However, an older university outside the group of research-intensive universities was just

beginning to "unpick" the global rankings. The corporate plan now referred to them and included a general 'aspiration to be in the top 200'. Nevertheless, it remained more concerned with establishing itself in the top 20 of *The Times* national table in order to consolidate its international student recruitment.

Following these early stages of coming to terms with rankings, institutions began to translate their newfound intelligence into strategic actions. Differences of emphasis between the types of institution were predictable: entry requirements and the correct assignment of research publications and Nobel Prize winners were a priority for the top-ranking institutions; graduate employment was significant for the smaller older university; and the NSS was more important for the former polytechnics and university college. There were more subtle differences, however, in the overall approaches. Those outside, but aspiring to the top echelons focused on tackling weaknesses and no longer tolerating poor performance, and were more willing to make resource allocations in an attempt to reach the position they felt they deserved. The highly ranked university focused more on what a "top-10 university should be doing," developing and refining its brand, and ensuring that its academics, alumni, partners, etc., "spread the word" to the key markets – in short, it was more concerned with communicating its ranking successes.

11.7.4 Affective Responses

The case studies provided persuasive evidence of the effect of rankings on the "collective psyche" of an institution and the strong emotional responses they can provoke, despite a common skepticism about the purposes of the publishers and the methods of the compilers. The general tenor of affective responses was predictable; the more highly ranked institutions found "solace" in their success as represented by their ranking position, which gave the staff a "boost" and helped students and alumni to "feel good" about their university. However, even here there were subtle distinctions between younger staff who were "thrilled to bits and felt part of the success" and older colleagues who were more "skeptical" about rises in ranking positions.

In all those institutions outside the upper echelons of the league tables, however, the predominant emotion was "hurt" – a deeply personal but also collective pain. Staff morale could be damaged by a poor result, especially if it occurred in an area that an institution believed it had focused on and developed strengths, such as supporting students' learning or preparing graduates for work. In such circumstances, there was "a sense of affront" and even "moral outrage" at the ways in which particular tables were perceived to have misrepresented the institution and the work of its staff. For example, in one former polytechnic, academics' professional qualifications and experiences as practitioners were thought to be devalued by a particular ranking system that gave greater value to academic research degrees. Staff felt aggrieved at times, particularly on behalf of the students, and especially when many of these were from disadvantaged families. Some staff felt it was almost as if the

publishers of the tables wanted to "destroy" the former polytechnics. A few wanted to write to the media to complain, but were advised not to by their institution because "it would look desperate."

League tables were described as a source of stress and as leading, in some cases, to individual distress. Poor results could produce considerable soul-searching, "a sense of impotence" and the feeling that "you can never do well enough." They provoked blame – of others within the institution and of oneself. They could even result in "a bit of madness," so that some staff were aggravated enough to ask why the institution should not "cheat" in its submission of data, to counterbalance the biases of the compilers. Anxiety was created by changes in ranking methodologies and the uncertainty about what other institutions might be doing to improve their standings. Longer term, these effects could undermine the self-esteem of staff and students, creating "a malaise that lingers" and a lasting "gloom" rather than "dramatic slumps in morale."

Perversely, some interviewees also felt that rank position could become almost an "addiction" and an institution could become "transfixed" by the tables. Being in "the wrong quartile" could produce "schizophrenia," especially in developing partnerships abroad. While denying the relevance of rankings, staff would have to acknowledge that potential partners used them, and yet, these same staff would themselves refer to the rankings when seeking new partners in unfamiliar territories or evaluating the competitors in new markets.

11.7.5 Self-management

All case study institutions had sought to optimize their data submissions. Few admitted to manipulating data, but some acknowledged it was possible to exploit ambiguity in a number of the definitions without actually committing falsification. Ranking data, and particularly the individual indicators that make up the league tables, were used as institutional key performance indicators and routinely included in management information systems, along with other data. Some of these data also featured on the national funding councils' Teaching Quality Information website, *Unistats*, and were required by the Quality Assurance Agency for Institutional Audit purposes, creating an interconnection between public accountability, quality assurance, and commercially compiled league tables. Institutional personnel frequently referred to the ways in which league tables gave additional weight to some of the individual indicators included, such as the NSS results and the proportion of good degrees awarded.

Both senior management and departmental staff described rankings as a lever for internal institutional change. On the one hand, they had helped senior administrators to gain backing from the vice-chancellor for actions that would not have been taken otherwise. On the other hand, senior academic managers admitted to using lower than expected results to put pressure on middle managers – or empower them – to introduce improvements. Deans and heads of departments and administrative units

had been "hauled in" by senior management to account for a poor showing. Peer pressure from other middle managers had "shamed" their colleagues into making amends or best practice had been disseminated from highly placed departments to those ranked lower. Disappointment with ranking positions had encouraged reviews and self-assessments, and benchmarking with other departments within the institution and with similar departments in other HEIs.

As well as feeding competition between institutions, league tables had also helped to foster competition between departments in the same university. This was encouraged either through the subject based tables produced by several of the publishers or by the institution itself disaggregating, by department, the data collated and submitted to the national agency and circulating them internally. Together with benchmarking exercises, this element of internal competition was given impetus by performance management systems that targeted areas of "weakness" and aimed to "raise the average" through all individual units improving their performance. Ranking systems had given "more focus" and attached "greater urgency" to such initiatives.

Two examples illustrate the longer term impact of these developments. First, several of the lower ranked case study institutions had conducted reviews of the curriculum, especially when planning new programs. These had included an assessment of the impact on league table indicators of establishing new programs and closing existing courses. In a number of cases, the decision was taken not to proceed with a new course that would lower entry requirements, reduce completion rates, or depress student survey results because the new students (for example, in particular vocational disciplines) were likely to be more critical than most. Second, senior management reported referring to the league tables before undertaking "headhunting exercises" to recruit senior academics. The implication was that candidates would only be approached if their current employer were a university ranked similarly to (or even higher than) the recruiting institution.

These examples, and others from the case studies and survey, suggest that rankings are constraining decision-making and channeling it in certain directions. In particular, they appear to generate among personnel in institutions a fear of failure, of criticism from peers and the media and of "blaming and shaming." Several interviewees highlighted the danger of this tendency engendering risk aversion and restricting innovation. One respondent from a former polytechnic claimed this was preventing the institution from becoming a "properly modern university."

11.7.6 Degrees of Control: Resisting, Managing, Exploiting, and "Gaming" the Rankings

A very small number of institutions in the UK have sought to mitigate the negative effects of league tables on their reputation by refusing to give permission to the national statistics agency, HESA, to release their data to compilers. There is

evidence that this number is growing, but it remains a tiny fraction, and the vast majority appear to wish to gain any kudos they can from whichever table or individual indicator shows them in a good light, or to avoid the imputed criticism of being afraid of being "shown up." In response to the threats to withdraw, compilers maintain they will simply substitute HESA data with information already in the public domain over which the institution will have no control.

Interviewees from all the case study institutions asserted that they were not "driven" by the league tables and some professed the naïve belief that focusing on the "right things" should automatically lead to improved ranking positions. Some respondents distinguished between "real quality" and the attributes favored by league table compilers. While they acknowledged that their competitors were almost certainly attempting to improve their own positions, there was no strong sense of the zero sum nature of rankings systems or the realization that they may have to expend a lot of effort just to "stand still" and maintain the same rank. Also, it was clear that the identification of the "right things" to concentrate on and what to do about them was being shaped by rankings systems and the key indicators employed by compilers. Those institutions developing more sophisticated approaches to rankings had at least identified which indicators they could have some impact on, applying "the 80:20 principle" (focusing 80% of their efforts on the 20% of the indicators they believed could be influenced). In the lower-ranked institutions, the focus tended to be on spending on facilities and "the student experience."

In the majority of cases, institutions had concluded they could do something about their student survey and first destination results (FDR), despite the lack of evidence for this. Many had mobilized final year students to complete the NSS on the assumption that "satisfied customers" tend not to respond as willingly as those who were dissatisfied, and so results would improve. Some had tried this with the FDR survey, but had only succeeded in slightly increasing the proportion of unemployed graduates recorded. A few lower-ranked institutions had taken a more differentiated approach to improving their NSS response rate by avoiding large departments with a record of being dissatisfied. These same institutions were also seeking to counter bad publicity circulating on social networking sites and to disseminate positive messages about their institution. One interviewee believed that a personalized and individualized approach to students at all stages of their relationship with the university, from applicant to alumnus, might be one way of circumventing the league tables in the future.

A number of "aspirant" institutions were troubled by the apparent tension or even conflict between government-supported initiatives to widen participation in higher education to disadvantaged groups with lower educational achievements and the pressure from the league tables to maximize the entry qualifications of their intake. The former polytechnics were particularly concerned about the effect of this conflict on local partnerships with further education colleges from where many of these non-traditional students transferred on linked programs of study. They wanted to influence compilers and to educate them on the importance of diversity and the greater usefulness of subject tables than institutional agglomerations. But they were frustrated at not being able to intervene with them. One older

university, however, had found a way around this by ensuring that these students remained registered with the college and would not be recorded as belonging to the university for HESA's (and therefore league tables') purposes. This had been agreed as legitimate by the national funding council. In compensation for this, the university was considering charging the college registered students lower tuition fees than its "own" students.

Senior managers at these same institutions acknowledged having to spend considerable amounts of time managing reactions to league table results and demonstrating – to governors, staff, and external constituencies – that they were taking an informed, professional, and realistic approach. They were trying to maintain a degree of "stability" and agree a level-headed and consistent attitude, "toning down" extreme reactions. They sought to "de-sensitize" the league table "issue" in the institution by "routinizing" and "accepting them." These managers would place their analyses of rankings in a wider context, provide a "filter swallow" and "spread some jam" around their reports.

For the highly ranked university, it was important that it maintained a consistently high ranking in most of the influential tables, despite their differing methodologies. This "showed strength in depth" at the university, because it was good at both teaching and research. Having a ranking that only focused on teaching might allow institutions specializing in this to predominate. But few universities could sustain a high reputation in both, and across a range of disciplines. As long as this university achieved high rankings, it could claim it was not driven by league tables but it had, nevertheless, identified a drop in ranking positions as a major risk. This would have an immediate impact, particularly on the university's external image, and especially in its international markets.

Finally, interviewees from several of the case study institutions were concerned about league tables becoming more of an issue in the future. They cited a downturn in the number of school leavers in the UK from 2012 reducing demand for higher education study, the possibility of the government allowing tuition fees to increase sufficiently to introduce variability between institutions and between subjects, and greater competition from private providers. These and other developments were thought likely to increase competition for students and enhance the influence of commercial rankings.

11.8 Discussion

This re-analysis of the six case studies included in the original study provides ample evidence and numerous examples of the ways in which different higher education institutions have been affected by, responded to, and used rankings at various points and at different levels of the organization. Despite the relatively small number of cases and the limited scope of the fieldwork and survey, this analysis illustrates how institutions at various positions in the rankings, operating in different markets and with contrasting histories, resources and reputations will differ in their

approaches to mitigating the negative effects and maximizing the advantages of rankings. Whether it is a "top" university seeking to sustain its reputation and improve its brand recognition globally, or a low-ranked institution "waking up" to the importance of reputation, learning the rules of the league tables game and "catching up" with its peers, rankings had exerted a major influence on institutional behavior. Clearly, the case study institutions were evolving in their responses and, no doubt, an institution's approach could shift significantly, for example, due to a change of leadership or of mission. Their tactics will surely continue to develop, not least as the methodologies of ranking systems are revised or the bases of particular indicators – such as the NSS or research quality assessment – are reformed. This penultimate section of the chapter discusses the foregoing analysis in the light of the conceptual frameworks presented earlier in the chapter and the developing marketization of higher education in the UK.

Clearly, rankings are reactive measures, as higher education institutions and their members are changing and being transformed by the ways in which the ranking systems evaluate institutional reputation and resources. The quantitative indicators selected by the compilers of UK league tables are largely those that are available rather than close proxies of the qualities they seek to represent, i.e., they count what is measured rather than measure what counts. They exclude much of what might be considered to indicate good quality or high performance because they reduce complex qualitative processes to shared metrics. So, for example, admissions processes become driven by the need to attract the highest qualified applicants; learning and teaching are reduced to "the student experience" and the ratio of staff to students; and careers services are steered to concentrate on immediate post-graduation employment.

Ranking systems also generate self-fulfilling prophecies. They employ a deficit model of a university that seeks to quantify the degree of inferiority to Oxford and Cambridge in the UK league tables and to Harvard in the international rankings (Little and Locke 2008). They do this by giving the "best" institution in the aggregated measures a maximum score of, say, 100 and calculating the lower scores according to how close they are to this maximum. This deficit model encourages lower status institutions to imitate those with high status by attempting to maximize their scores in the key indicators. This leads to isomorphism among higher education institutions and undermines diversity within the national system.

The transformation of higher education institutions by rankings within an increasingly marketized environment occurs initially through a process of internalization of ranking systems logic by organizational members who are then seduced and compelled to institutionalize this in processes, systems, and structures. Despite – or, perhaps, because of – an initial sense of dissonance between the actual and expected (or desired) ranking position, institutional members seek to better understand the ranking methods and how their institution's data contribute to its relative positioning. On discovering they cannot (except for the influential "elite" institutions) persuade the compilers to modify their ranking systems to fit their own institutional model and mission, they find ways to optimize their data to fit the existing rankings. If the institution's trajectory is upward in the tables, organizational

members will feel good about this external recognition, despite their enduring skepticism. If it is downward, they will feel devalued and demoralized and seek to offer excuses and justifications and criticize the publishers of league tables but, nevertheless, remain obsessed by the rankings. Either way, institutions will assert they are not driven by league tables while doing their utmost to keep up with – or ahead of – their peers and managing the perceptions and expectations of their key stakeholders. Every decision then needs to be assessed for its likely impact on the institution's ranking position. Ranking systems logic becomes normalized, and hence legitimized, if reluctantly. Gradually, and subtly, this begins to change perceptions of higher education, expectations of institutions and the behavior of their members.

Having identified those elements of league tables that the organization might have some influence over – student survey responses or expenditure on library and computing facilities, for example – institutions seek to make changes. Efforts may be made to introduce benchmarking and "peer competition," to bring all units up to the level of the best performing departments, and to lever institutional change. League table measures are introduced into management information systems and the data are disaggregated by department, unit or function. Resources may be redistributed or partially allocated in ways that are "ranking friendly." Work is redefined and becomes more "customer-focused" and "business-facing." Programs are reviewed, new partners are assessed and recruitment is informed by reference to the rankings. Almost regardless of the position of the institution in the tables, the marketing professionals will find some way of using rankings to promote the organization to its major markets, even if this means being highly selective, only comparing the institution with a limited range of "peers," or constructing entirely new tables to show it in a favorable light. Increasingly, institutions are adopting improved ranking positions as an explicit institutional goal. By these, and other, means, the logic of ranking systems becomes embedded in institutional practices, routines, plans, and, ultimately, missions.

However, as this chapter has sought to demonstrate, these processes of internalization and institutionalization vary between types of institutions at different places in the rankings. Those universities who expect to be at, or near, the top of the tables seem to respond more quickly and more strategically to rankings, and are primarily concerned with exploiting the reputational value of their ranking success and using this to position themselves centrally within their key markets, largely through initiatives designed to increase brand recognition. Those just outside the research-intensive group tend to focus first on the "frontline" activities of student recruitment and marketing and on efforts to improve their standing in the national tables. It is likely, however, that they will begin to target the world rankings, as pressures on the domestic student market for students, academics and research funding encourage them to develop an increasingly global brand.

Higher education institutions in the bottom half of the national tables were initially more concerned with "catching up" and ensuring they were doing at least what every other HEI was (assumed to be) doing to "optimize" their data submissions. Subsequently, they focused on those indicators they believed they could influence, such as the NSS, entry scores, retention rates, and "good degrees

awarded." More recently, they had begun to review their curriculum and to carefully scrutinize new program proposals for their likely impact on their league table positions. Those excluded from the main institutional tables (although not all subject tables) lobbied for their inclusion in them.

Responses to rankings may be more similar between different types of higher education institution to start with but, as institutions become more sophisticated in their approaches, and as small differences between them become magnified and exaggerated (and even created) by ranking systems, their strategies gradually become differentiated by their positions in clusters of institutions with similar ranking positions. They become ensnared by different self-fulfilling prophecies according to whether they feature in the world rankings or the national tables, and which indicators they perceive they can improve on in the overall national tables. Accordingly, they may decide, for example, to bolster their global reputations by concentrating resources on highly cited researchers in science fields, modify their curricula to maximize graduate employment, or emphasize how student-focused they are in providing an "excellent" learning experience.

11.9 Conclusions

Rankings have both facilitated and shaped the marketization of higher education in England, the UK as a whole, and elsewhere. They have facilitated marketization by introducing greater competition between and within higher education institutions. Ultimately, they accomplish the transformation of qualities into quantities, which is both required by, and a consequence of, the commodification and privatization of higher education. Rankings have also helped to embed the logic of the market within organizational structures and processes and within the minds and practices of organizational members. They influence institutions to become more business-like (Martins 2005). They have enabled senior institutional managers to foster internal competition between academic units and create internal markets. In some ways, in a highly regulated UK higher education market (Locke 2010), rankings have become a substitute for more authentic market mechanisms, such as student vouchers, unregulated tuition fees, and the free market entry of private providers of higher education services.

However, UK higher education continues to be dominated by an enduring reputational hierarchy of institutions and, of course, ranking systems are sustained by, and themselves, reinforce this hierarchy (while, at the same time, modifying it). Competition between institutions is localized within the rankings, occurring primarily between those of a similar ranking position, and the nature of this competition varies at different points in the rankings (Grewal et al. 2008). So, the efforts of highly ranked universities and lower placed institutions to improve their reputation and increase the resources available to them are very different. And, while the compilers of the national rankings – along with governments – try to organize all higher education institutions into a single system, in reality, different types of

institutions operate in very different markets. Ranking systems also significantly modify and reshape higher education markets by appearing to influence institutions' major "customers" and external constituencies, such as prospective students (domestic and international), employers of graduates, "lay" governors, governments and their intermediary agencies, and research funders. By doing so, they create new forms of inequality between institutions (Sauder and Lancaster 2006).

The empirical evidence and analysis presented in this chapter clearly indicates the need to go beyond the investigation of "impacts" and develop an understanding of how higher education institutions start – and continue – to engage with processes of marketization, as a way of surviving, prospering and managing status anxiety in changing and challenging environments, and how this is made possible and modified by ranking systems. This suggests an agenda for further in-depth investigations of the forms of internalization and, ultimately, the institutionalization of ranking system logics and processes by organizational members. The evidence and interpretation offered here also indicates the need for these investigations to carefully differentiate between:

- The levels of an institution and their different responses to, and uses of, rankings
- Disciplines or fields of study (for example, how law and business schools compare with medical schools and science disciplines)
- Types of higher education institutions, the kinds of ranking systems and individual measures they regard as important, and the particular indicators they perceive they can improve their performance in. Also relevant are the resources at their disposal to mitigate the effects or exploit the influence of rankings
- The stage at which an institution is responding to rankings – both national and global – and, therefore, the degree to which the processes of internalization and institutionalization have taken place

This chapter has sought to understand the influence of ranking systems in a particular higher education market. Clearly, ranking systems are not markets, and markets may not require rankings in order to operate successfully. But the interactions between rankings and markets go some way to explaining why rankings have become so influential. Ultimately, judgments about the validity of ranking systems are not as important as the influence they exert on – and within – institutions (Martins 2005).

> In some significant sense, all the things wrong with the rankings matter considerably less than the plain fact that the rankings matter. (Gioia and Corley 2002: 112)

This is what now makes rankings difficult to resist, let alone boycott altogether.

References

Bastedo, M., & Bowman, N. (2011) College rankings as an interorganizational dependency: establishing the foundation for strategic and institutional accounts, *Research in Higher Education, 52*(1), 3–23.

Bowman, N., & Bastedo, M. (2009). Getting on the front page: Organizational reputation, status signals, and the impact of U.S. news and world report on student decisions. *Research in Higher Education, 50*(5), 415–436.

Clarke, M. (2007). The impact of higher education rankings on student access, choice, and opportunity. *Higher Education in Europe, 32*(1), 59–70.

Cyrenne, P., & Grant, H. (2009). University decision making and prestige: An empirical study. *Economics of Education Review, 28*(2), 237–248.

Elsbach, K. D., & Kramer, R. M. (1996). Members' responses to organizational identity threats: Encountering and countering the business week rankings. *Administrative Science Quarterly, 41*(3), 442–476.

Espeland, W. N., & Sauder, M. (2007). Rankings and reactivity: How public measures recreate social worlds. *The American Journal of Sociology, 113*(1), 1–40.

Gioia, D. A., & Corley, K. G. (2002). Being good versus looking good: Business school rankings and the Circean transformation from substance to image. *Academy of Management Learning and Education, 1*(1), 107–120.

Grewal, R., Dearden, J. A., & Lilien, G. L. (2008). The university rankings game: Modeling the competition among universities for ranking. *American Statistician, 62*(3), 232–237.

Hazelkorn, E. (2007). The impact of league tables and ranking systems on higher education decision-making. *Higher Education Management and Policy, 19*(2), 81–105.

Hazelkorn, E. (2008). Learning to live with league tables and ranking: The experience of institutional leaders. *Higher Education Policy, 21*(2), 193–216.

Hazelkorn, E. (2009). Rankings and the battle for world-class excellence: Institutional strategies and policy choices. *Higher Education Management and Policy, 21*(1), 55–76.

IHEP. (2009). *Impact of college rankings on institutional decision-making: Four country case studies, issue brief*. Washington, DC: Institute for Higher Education Policy.

King, R. (2009). *Governing universities globally: Organizations, regulation and rankings*. Cheltenham: Edward Elgar.

Little, B., & Locke, W. (2008, September 11–13). *Conceptions of excellence in teaching and learning and implications for future policy and practice*. Paper presented at the Consortium of Higher Education Researchers (CHER) 21st Annual Conference, Excellence and Diversity in Higher Education. Meanings, Goals and Instruments, Universita degli Studi di Pavia, Italy.

Locke, W. (2008, December 9–11). *A "league table culture"? The policy and governance implications of university rankings in a hierarchical higher education system*. Paper to SRHE Annual Conference, Liverpool.

Locke, W. (2010). False economy? Multiple markets, reputational hierarchy and incremental policymaking in UK higher education. In R. Brown (Ed.), *HE and the market*. New York: Routledge.

Locke, W., & Botas, P. (2009, September 10–12). *The academic labour market in the United Kingdom: Fully developed or quasi-market?*. Paper presented at 22nd Annual Conference of the Consortium of Higher Education Researchers, Porto.

Locke, W. et al. (2008). *Counting what is measured or measuring what counts? League tables and their impact on higher education institutions in England* (Report to HEFCE). Bristol: Higher Education Funding Council for England. http://www.hefce.ac.uk/pubs/hefce/2008/0814/

Marginson, S. (2007). Global university rankings: Implications in general and for Australia. *Journal of Higher Education Policy and Management, 29*(2), 131–142.

Marginson, S. (2009). The knowledge economy and higher education: A system for regulating the value of knowledge. *Higher Education Management and Policy, 21*(1), 39–53.

Marginson, S., & van der Wende, M. (2006). *To rank or to be ranked: The impact of global rankings in higher education*. Enchede: Centre for Higher Education Policy Studies, University of Twente. http://www.studiekeuzeenranking.leidenuniv.nl/content_docs/paper_marginson_van_der_wende.pdf

Marginson, S., & van der Wende, M. (2007). *Globalisation and Higher Education* (Education Working Paper No. 9). Paris: Organisation for Economic Co-operation and Development, especially Chapter 8 'University rankings and typologies': 55–62. http://www.oecd.org/dataoecd/33/12/38918635.pdf.

Martins, L. L. (2005). A model of the effects of reputational rankings on organizational change. *Organization Science, 16*(6), 701–720.

Meredith, M. (2004). Why do universities compete in the ratings game? An empirical analysis of the effects of the "U.S. News and World Report" college rankings. *Research in Higher Education, 45*(5), 443–461.

Mohrman, K., Ma, W., & Baker, D. (2008). The research university in transition: the emerging global model. *Higher Education Policy, 21*, 5–27.

Monks, J., & Ehrenberg, R. G. (1999). *The impact of "U.S. News & World Report" college rankings on admissions outcomes and pricing policies at selective private institutions* (Working Paper 7227) Cambridge: National Bureau of Economic Research (NBER). http:// www.nber.org/papers/w7227.

Morley, L., & Aynsley, S. (2007). Employers, quality and standards in higher education: shared values and vocabularies or elitism and inequalities? *Higher Education Quarterly, 61*(3), 229–249.

Salmi, J., & Saroyan, A. (2007). League tables as policy instruments: Uses and misuses. *Higher Education Management and Policy, 19*(2), 1–38.

Sauder, M., & Espeland, W. N. (2009). The discipline of rankings: Tight coupling and organizational change. *American Sociological Review, 74*(1), 63–82.

Sauder, M., & Lancaster, R. (2006). Do rankings matter? The effects of U.S. news and world report rankings on the admissions process of law schools. *Law & Society Review, 40*(1), 105–133.

Teichler, U. (2008). Diversification? Trends and explanations of the shape and size of higher education. *Higher Education, 56*, 349–379.

Walpole, M.-B. (1998, April 13–17). *National rankings: Ramifications for academic units*. Paper presented to the annual meeting of the American Educational Research Association, San Diego.

Wedlin, L. (2006). *Ranking business schools: Forming fields, identities and boundaries in international management education*. Cheltenham: Edward Elgar.

Chapter 12
Reaction to Academic Ranking: Knowledge Production, Faculty Productivity from an International Perspective

Akira Arimoto

12.1 Introduction

Considering the fact that academic ranking has been introduced internationally by agencies such as the London Times (THES), Shanghai Jiao Tong University, and US News and World Report, we can predict an institutionalization of academic ranking throughout the world. Almost all countries are reacting to such situations by competing for high rankings for their institutions. Yet, the results of the rankings attract much criticism because of the inadequate criteria they employ.

On the other hand, it is also natural for every government, university, academic, and other individuals to pay attention to rankings, since they are apt to have a global impact. In fact, there is tendency for governments and universities to overrespond to the rankings by seeking to strengthen systems and institutions in readiness for competition.

In this context, the academic profession should pay attention to the kind of response that is being made. It is the academics as agents, or main actors, who are directly committed to the academic work of research, teaching, and service that usually provide indicators for the process of academic ranking, and who contribute to the development of academic productivity through both research and teaching productivity.

This chapter seeks to deal with the main theme "Reaction of the Academic Profession to Academic Ranking," analyzing it from the author's own perspective. The theme of "Reaction of the Academic Profession to Academic Ranking from an International Perspective," is approached with a focus on the USA and Japan as case studies. The main materials used for the following analysis are based on various sources, including the author's proceeding articles, and the results of the CAP (Changing Academic Profession) survey, which was conducted in 2007 by 18 countries (Arimoto 2008, 2009a, b; RIHE 2009).

A. Arimoto (✉)
Research Institute for Higher Education, Hijiyama University, Hiroshima, Japan
e-mail: akira.arimoto@ksu.ac.jp

J.C. Shin et al. (eds.), *University Rankings*, The Changing Academy – The Changing Academic Profession in International Comparative Perspective 3,
DOI 10.1007/978-94-007-1116-7_12, © Springer Science+Business Media B.V. 2011

12.2 Knowledge Production: Framework of Research

12.2.1 Paradigm Shifts of Knowledge Production

As shown in Fig. 12.1, the academic profession as well as academia (universities and colleges) is changing in accordance with social changes. The essence of the structure in the figure means that academia is changing from "academia 1" (A1) to "academia 2" (A2) in accordance with the environmental changes. The first great environmental change is a social change caused internationally by trends such as globalization, knowledge-society orientation, marketization, etc. Through similar trends, modern universities were established in industrial society, and A1 structure was developed at the time of the post-industrial society and the information society. An A2 structure, emerging today as a knowledge society, develops from an information-orientation society. This has been apparent from the 1960s. The concept of the knowledge society, which Peter Drucker argued for in his "Post-capitalist Society" (Drucker 1993), has spread over the world since then so that the universities and colleges have been forced to respond to this environmental change. At the time of A1 structure a community of knowledge was prevailing, while in the A2 age an enterprise of knowledge is increasingly prevailing, in which an entrepreneurial university and even academic capitalism are appearing (Clark 1998; Slaughter and Leslie 1997).

The other large environmental change is a reconstruction of knowledge. As Michael Gibbons et al. pointed out in 1994, reconstruction of knowledge from

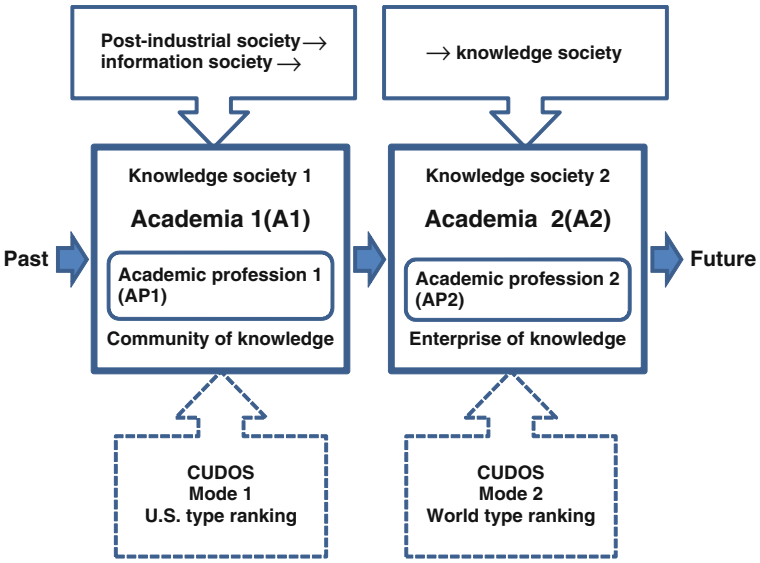

Fig. 12.1 Framework of research

"mode 1" to "mode 2" has occurred (Gibbons et al. 1994). Taking this into account, the response of traditional academia has to be a transformation to a new academia structured for "mode 2." Academia had already constituted a knowledge society in prior years, because it ceaselessly pursued academic work on the basis of academic disciplines, or advanced knowledge. But it is considered a knowledge society corresponding to knowledge of "mode 1" type. Today, total society has become a knowledge society incorporating knowledge of both "mode 1" and "mode 2" types. In this context, A1 needs to be transformed to a new A2 to embrace not only a knowledge society but knowledge itself.

Given this, academic staff has to change from being "mere academic staff" to becoming an academic profession, or strictly speaking, a status of achieving academic professionalism. Specifically, as described below, it was in the second half of nineteenth century when the graduate school was institutionalized in the USA and a Ph.D. was established as a "union card" to enter academia. As Light (1974) described it, the academic profession emerged when the graduate school worked as a function of preparation for an academic career. The academic profession at the time of a community of knowledge is considered to be "academic profession 1," while the counterpart in the era of an enterprise of knowledge is considered to be "academic profession 2." The former is involved in knowledge of "mode 1" type and an ethos of CUDOS type; the latter is involved in knowledge of "mode 2" type and an ethos of post-CUDOS type. Robert Merton (1973) named CUDOS and it is mostly adaptable to "academism sciences" in the traditional academia (Merton 1973; Arimoto 1987).

Summarizing the above, we can recognize the structural development from A1 to A2 (Table 12.1) as follows: the social change = transformation from post-industrial society to knowledge society; value = from unified to diversified; knowledge = from "mode 1" to "mode 1" plus "mode 2"; ethos = from CUDOS to post-CUDOS; the enterprise = community of knowledge to enterprise of knowledge; academic work = research paradigm to reconstruction of knowledge; the academic profession = from particularism to universalism; the reward system = from ascription to achievement; the academic ranking = from the US type to the International type.

The academic ranking which this chapter discusses in the given structure of the framework can be separated into two stages: The first stage is A1, in which the

Table 12.1 Structure of A1 and A2

Criteria	A1	A2
Social change	Post-industrial society	Knowledge society
Value	Unified	Diversified
Knowledge	Mode 1	Mode 2
Ethos	CUDOS	Post-CUDOS
Organization	Community of knowledge	Enterprise of knowledge
Academic profession	Particularism	Universalism
Academic work	Research paradigm	Scholarship reconsidered
Ranking	US type	Worldwide type

ranking was undertaken nationally in the academic marketplace in the USA, that started in the second half of the twentieth century and stressed research orientation and productivity for the first time. The second stage is A2, in which the academic ranking was undertaken internationally in a global academic marketplace from around 2003. These two stages have imposed differences to the point that the ranking in the first stage mostly affected academics in the USA and stimulated their reactions to it, while the ranking in the second stage affected academics elsewhere. The academic professions worldwide, who comprise "the small worlds, different worlds" Burton Clark identified would be expected to respond to the academic rankings differently according to the systems, cultures, climates, and disciplines to which they belong (Clark 1987).

12.2.2 Knowledge Production and Research University

Knowledge is conceived to be quite important to the study of higher education from the perspective and methodology of the sociology of science (Merton 1973; Becher 1981, 1989; Clark 1983, 1995, 2008; Jacob and Hellstrom 2000; Becher and Parry 2007; Bleiklie and Henkel 2005; Kogan et al. 2006; Parry 2007; Arimoto 1981, 1987, 2007). The knowledge function consists of discovery, dissemination, application, and control, and in other words it means research, teaching, service, and administration, and management, respectively. Among these functions, research, teaching, and service make up academic work, which is identified as the most important function in the activities of universities and colleges (Clark 1983; Arimoto 1981, 2005). In academic work, research and teaching are thought to be the main functions for the academic profession. In particular, academics are engaged in "academic productivity" which Michiya Shinbori distinguished from "scientific productivity" as used by Robert Merton (Merton 1973; Shinbori 1973). Academic productivity mainly consists of research productivity and teaching productivity (Arimoto 1981, 2005).

Academic productivity is perceived to have a logic of connecting itself with the concept of "center of learning" (COL), or "center of excellence" (COE), when it has high visibility in terms of quantity as well as quality (Ben-David 1977; Arimoto 1996). In fact, universities attracting many prominent academics with high productivity form centers of learning. Other distinguished researchers and teachers tend to move to these centers, if possible, once they have gained recognition.

Paris and Bologna were universities with characteristics of COE in the early Middle Ages because of their reputation for teaching productivity rather than research productivity. The mobility of teachers and students toward the COE institutions at that time, when no indicators and methodologies existed to evaluate academic productivity, was thought to be in response to the reputation of the Universities. On the other hand, in the COE institutions in modern universities, initially found in German universities where research was institutionalized, the reputation for research productivity has become more important than teaching

productivity. As a result, it is there we can observe the beginnings of a research paradigm and also the appearance of a research marketplace mechanism.

Following the establishment of Berlin University in 1810 and the institutionalization of scientific knowledge in German universities, the "research paradigm" emerged, with a focus on discovery of knowledge and the relationship of the COE conforming to research productivity.

A comparative study of "eponymy" testifies that the COEs were in Germany in the nineteenth century (Ruffner 1977; Shinbori 1985; Arimoto 1996). The COEs were in France and the UK in seventeenth to eighteenth centuries, in Germany in the nineteenth century, and in the USA in the twentieth century. Many researchers and students were sent to these centers from all over the world, returning with many experiences of culture, academic climate, and ethos to their native countries. As a result, research productivity became a greater priority internationally as well as nationally. This research orientation and productivity went from Germany to the USA where it has subsequently been encouraged to a great degree.

During the nineteenth century, approximately 8,000 students were sent to German universities from the USA (Oleson and Voss 1979). Some graduate schools such as Johns Hopkins, Clark, Harvard, Yale, and Princeton that were established in the second half of the nineteenth century were involved in importing the German style of research orientation. They promoted competition for German-type academics by paying considerable attention to scientists, scholars, and researchers with outstanding research productivity. Preeminent universities such as Harvard, Yale, and Princeton started to change from colleges for liberal arts education to universities for research and professional education by recruiting researchers holding doctoral degrees (Pierson 1952).

Some other countries including Japan sent their own students to the centers of learning throughout the world, especially Germany, importing a model of research orientated to their own countries (Nakayama 1978).

As described above, both the US universities and the academics reacted swiftly to catch up with the level of academic productivity of the German universities, and engaged in the series of academic reforms described below.

12.2.2.1 Curriculum Reforms

The introduction of an elective system by Charles W. Elliot, President of Harvard University, promoted a transformation from the old style of "recitation" to a new style of "teaching through research." It was almost the same at Yale and George Pierson described it as follows: "Many criticized the old college curriculum because it was too narrow or elementary. Especially, the men who had been trained in Germany wanted to introduce the German ideals of free teaching and free study, of lectures rather than recitations, and of specialized investigation and research" (Pierson 1952: 45). It was an epoch making event in American higher education resulting in the fact that research has become the basis of teaching.

12.2.2.2 Restraint of Inbreeding in Harvard University and Yale University

Universities such as Johns Hopkins and other institutions with high research productivity no longer recruited their own graduates but Ph.D. holders as academic staff. Pierson described the situation at Yale: "In 1908, the Alumni Weekly printed a statistical summary which revealed that from 1801 to 1877, Yale College has appointed only one non-Yale man to a professorship. From 1877 to 1900, a broader policy had introduced a nucleus of outside talent, without sensibly threatening the established order. And since 1900, more than half the professorial appointments and more than one-third of the junior appointments had gone to graduates of colleges other than Yale"(Pierson 1952: 291). This represents a dramatic paradigm shift from the teaching orientation, dating back to the universities of the Middle Ages, to the research orientation of the Modern universities.

The shift away from hiring their own graduates started from that time and has continued for more than a century until today. Harvard, which introduced the German research orientation model, became a pioneer in promoting the research paradigm. The efforts of conducting academic reforms have lasted for many years since then. An example is provided by Keller and Keller, who wrote in 1953, "Nearly half of the 448 members of the Faculty of Arts and Sciences had Harvard Ph.D." That the Harvard habit of promoting from within was declining was shown by the fact that "Of 68 senior appointments between 1953 and 1957, more than half came from other schools; only six had come up through the College, the Harvard Graduate School, and the Junior faculty. The inbreeding ratio at this time was controlled as low as 5% among all academic staff."(Keller and Keller 2001: 211).

The most important strategy for a research university to increase its competitiveness is to recruit researchers with high research productivity. Harvard is controlling its inbreeding ratio at a low level, even though it is the leading US institution producing the best and brightest graduates. It is understandable that the restraint of inbreeding is necessary to become a competitive institution in terms of research productivity. By contrast, the University of Tokyo, which is thought to be an equivalent to Harvard in terms of prestige among Japanese institutions, had high inbreeding ratios of 90% in 1965 and 80% in 2005 (Shinbori 1965; Arimoto 1981; Yamanoi 2007). This fact indicates that it has had no explicit philosophy of controlling inbreeding in the century since its establishment.

12.2.2.3 Institutionalization of the Graduate School

The German research orientation was realized by its transplantation to a graduate tier, which was newly established for the first time in higher education, replacing the undergraduate tier which had lasted for about eight centuries since the Middle Ages. The graduate school was established in Johns Hopkins University in 1876 as a base for research as well as for professional education, and other influential universities including Clark, Harvard, Yale, Princeton followed (Brubacher and Rudy 1968: 183). The Ph.D. became a sort of "union card" for entry to an academic

career (Veysey 1965: p. 176). Academic rank was introduced in Chicago University in 1891 in order to more effectively recruit and promote academics. Academic ranks consist of three grades and eleven classes: chief professor; professor; associate professor; assistant professor; instructor; associate; assistant; docent; lecturer; reader; and fellow (Rudolph 1962: p. 398). A new academic career was institutionalized to promote from fellow to professor under the pressure of competition. Research universities gradually emerged from these graduate schools and their research functions.

12.2.2.4 Some Reforms Such as Establishment of Academic Associations, Sabbatical Years, and University Presses

Academic associations coped well with the research orientation from inside and outside academia. A system of sabbatical years was started by Harvard in 1880 with a 7-year interval in order to increase research productivity. Academics were expected to publish books and articles after spending more or less than a year absent from teaching. In addition, a university press was founded by Johns Hopkins in 1881 and a series of university presses including Chicago, California, Princeton, Yale, and Harvard were founded to publish Ph.D. dissertations. As a result, a research orientation was further promoted (Rudolph 1962: p. 407).

Many other countries have attempted to respond to the German type of research orientation. Their efforts have not been as effective as that of the USA where many reforms were carried out to achieve the goal. Japan is not exceptional in its inadequate responses. Catching up with the level of science in other advanced countries has become a national policy in Japan for more than a century from the Meiji Restoration until today. The national government and academics have sought to evaluate the COE formed internationally in the different academic disciplines.

First, with the establishment of Tokyo Teikoku Daigaku (University of Tokyo), the academics tried to introduce the strongest disciplines and departments as developed in advanced countries. This is evidence of their fairly accurate insight into the centers of learning in the world at that time. For example, they selected the following disciplines from advanced countries as early as 1870: from the UK: mechanics, business methods, geology, architecture, and shipbuilding; from France: law, international law, biology, census, astronomy, physics, chemistry, and architecture; from the USA: mail systems, industrial arts, agriculture, farming, commerce, and mineralogy (Nakayama 1978: 42–43).

As Shigeru Nakayama pointed out, in the early Meiji years, the government and the academics decided to import scientific knowledge from the West, designating the advanced countries to invite prominent researchers and teachers. They sent students to these selected countries after asking them to study languages before leaving Japan (Nakayama 1978: 45).

Second, while they invited researchers and teachers from these advanced countries they did not introduce similar reforms to control inbreeding. Unlike the USA, they did not invite prominent scholars from other universities in Japan to participate

and thereby controlling the inbreeding ratio. Accordingly, it was inevitable that extension of prestige stratification led to an increase in the inbreeding ratio (Arimoto 1981; Yamanoi 2007).

Third, they sent many students overseas to the COEs. This is a sign that they sought to catch up with the scientific standards of the centers of learning in the advanced countries. However, at that time there was a great gap between the advanced countries and Japan as shown in the fact that Erwin von Bälz, who was invited in 1876 to Tokyo Medical School (University of Tokyo), wrote in his diary: "Japanese students tend to get ripe fruits from the tree of science instead of the spirit of its roots that produce them" (Bälz 1979).

At that time, students were still involved in "recitation" of materials taught by their teachers without thinking about them critically. This method is almost similar to that used in Harvard College in the early nineteenth century when student's recitation was popular as described by Pierson (1952: 45). This method had been in place for many years before an elective system was introduced into Harvard, and many innovative reforms in teaching and learning methods were not successful for a long time in the University of Tokyo and other institutions (Ushiogi 1984).

As shown above, both Japan and the USA paid attention to the German model of research orientation but Japan was not successful in establishing graduate schools as well as the universities to provide a research orientation. With hindsight, one could have predicted fairly easily more than a century before that Japan would remain well behind the USA in the future ranking order when the global institutionalization of academic rankings commenced in the twenty-first century

12.3 Academic Contexts of University Ranking

12.3.1 Ranking in the USA

From an international perspective, the first academic ranking was undertaken in the USA in 1925, when a simple form was used in the field of sociology, and again in 1960 when an improved one was introduced (Arimoto 1981: 132–138). There are clearly reasons for the initial institutionalization of ranking in the global academic community.

First, the research and science orientation had already been established before the introduction of academic ranking. The new emphasis on supporting a research orientation, which had been developed in German universities, was institutionalized in the academic community in the late nineteenth century, replacing the old value of emphasizing the teaching orientation that had been in place for almost six centuries since the Middle Ages.

Second, the institutionalization of the graduate school occurred separately from the undergraduate college because of its connection to a research orientation. The first step in this reform was made in 1876 by Johns Hopkins University, followed by Clark University in 1887, and the University of Chicago in 1892, which was

successful in forming both a graduate school and a research university. "The establishment of Johns Hopkins was perhaps the single, most decisive event in the history of learning in the Western hemisphere" (Shils 1979: 28).

Third, an academic departmental system was developed so as to promote academic productivity. Although it was originally developed on the basis of "departmentalism," in which control of an academic guild is established in the chair (Clark 1983), the basic unit of the research orientation in German universities was an institute in natural sciences and a seminar in humanities and social sciences on the basis of the chair system. The counterpart in the US universities was a department with a focus on a research orientation instead of the chair system which was not imported into the USA. As a result, it is thought that the department system is more likely to stimulate research productivity than the chair system (Clark 1983).

Fourth, based on these trends, the competition for pursuing quality assurance at the individual department level was promoted to the extent that culture and a climate for quality evaluation of academics was increasingly encouraged in the individual department. To define the quality of a department, various organizations and methodologies were invented at the same time: the organization for publication such as academic journals and the university press; the organization for assessing academic productivity such as academic associations; and the methodology for promoting publication in terms of economy and time such as sabbatical years (Rudolph 1962: 407).

Fifth, we can point to a social climate in the 1880s, when Henry Rowland made a comparison between Germany and the USA in physics, emphasizing research orientation. He proclaimed the need for construction of the best science institutions instead of "a cloud of mosquitoes" type of institutions. "Best science required an institutional pyramid, commanded at the heights by a best-science elite and open to talent at the bottom" (Clark 1983: 257).

Such reasons are thought to have worked well related to the appearance of academic ranking in the USA at an early stage of modern university history.

Thus, some significance can be attached to the fact that institutionalization of academic rankings was first undertaken in the USA.

First, the fact that ranking started about half a century ago reflects the operation of a market mechanism among institutions as well as individual academics so as to promote a priority competition for academic productivity. This anticipated the situation today when academic ranking is gradually extending over the global academic community accompanied by such priority competition.

Second, specific universities are apt to be situated at the top of the hierarchy. For example, in the case of departments of sociology in graduate schools, the top ten are as follows (Arimoto 1981: 136): in 1925, Chicago, Columbia, Wisconsin, Minnesota, Michigan, Harvard, Missouri, Pennsylvania, North Carolina, and Yale; in 1957, Harvard, Columbia, Chicago, Michigan, Cornell, Berkeley, Minnesota, North Carolina, Washington (Seattle), and Yale; in 1970, Berkeley, Harvard, Chicago, Michigan, Wisconsin, North Carolina, UCLA, Cornell, Johns Hopkins, North Western, and Princeton. In all the departments, including departments of sociology in the US university system, the research universities are apt to form the upper

stratum of the hierarchy. As argued by Parsons and Platt, the function of knowledge discovery has the highest prestige in academic work in American academia (Parsons and Platt 1973).

It has taken many years to create such a structure in the USA, since the establishment of Johns Hopkins University with its graduate school. As mentioned above, many universities, including Harvard, tried to reform their organizations by the recruitment of distinguished academics (and students), by the reinforcement of inbreeding, the establishment of departmentalism, and the institutionalization of a university press. Through these reforms, some institutions were gradually transformed from colleges to universities, eventually forming the top-ten group in the ranking order. For example, Morton and Phyllis Keller pointed out that Harvard with a score of 63 was placed at the top of the hierarchy in 1937 in their scores of the 28 fields of GSAS(Graduate School of Arts and Sciences) , followed by Chicago (121), Columbia (126), Yale (157), California (189), Johns Hopkins (199), Cornell (234), Princeton (242), Michigan (245), and Wisconsin (250) (Keller and Keller 2001: 110).

Through these processes, some were ranked by a science index as top-science institutions by Hugh Graham and Nancy Diamond who identified the top 20 leading public and private research I and II institutions in the 1990s (Graham and Diamond 1997: 109); public institutions: UC Berkeley, UC San Diego, Wisconsin-Madison, Colorado, SUNY-Stony Brook, Purdue, Illinois-Urbana, UCLA, Utah, and Arizona; private: Caltech, MIT, Rockefeller, Princeton, Stanford, Brandeis, Cornell, Harvard, Johns Hopkins, and Chicago.

It is important to acknowledge Vanervar Bush's efforts to create an American academic structure comparable to the German academic structure of the early twentieth century (Bush 1945). Much money was also invested in research universities by the US federal government and by private foundations (Geiger 1986, 2004). Responding to these trends, the research universities developed the US type of research orientation of about 200 institutions as defined by the Carnegie classification (Carnegie Commission on Higher Education 1976). The inclusion of about 200 research universities in a system makes it perhaps the largest one in the world, because in Japan, for example, it was approximately 25 institutions in 1980, when the Carnegie classification was adapted to the Japanese institutional situation (Amano 1984).

Third, both upward and downward mobility among institutions are recognizable. This is apparent in the previous data, in which, for example, Chicago declined from the top to a lower position, while Harvard moved up from sixth to the top position. Thus, examples of both "retention" and "scrap-and-build" have remained in the series of rankings.

Fourth, research universities, highly ranked in the hierarchy pyramid attract academics with high research productivity. Jonathan and Stephen Cole identified four types of academics based on a combination of their quantity and quality of research productivity: prolific (+ +), mass productive (+ −), perfectionist (− +), and silent (− −) (Cole and Cole 1973: 92). Jones, Lindzey, and Coggeshall undertook a survey in 1982, ranking the top 50 departments in various fields including

mathematics and physical sciences, biological sciences, engineering, social and behavioral sciences, and humanities (Jones et al. 1982). The author of this chapter conducted a survey in 1987 of 287 chairpersons who belonged to the top 50 departments ranked by Jones et al. The results indicate that while there are many academics of high quality productivity in high ranking departments, there are few in low ranking institutions. The 228 respondents reported that as many as 82.8% of their staff achieved high quality productivity. On the other hand, in non-high-ranking departments the proportion fell to 44.3%. It is also interesting to note that in the top ranking departments, the "silent type" constitutes 9.9%, and is less than the 35.7% in non-high ranking departments (Arimoto 1994: 27–47). In the top ranking institutions, academics pursue high productivity both in terms of quantity as well as quality.

Fifth, ranking is a reflection of the academic marketplace, which is working of a priority competition for academic productivity among institutions and academics internalized an ethos of science and research orientation. In other countries such as Japan where a market mechanism is working less positively and "sponsored mobility" is working rather than "contest mobility," an environment of academic ranking has barely been developed.

Sixth, a principle of meritocracy which emphasizes academics' research achievement is operating, in addition to evaluation and reward systems which also emphasize a research orientation. In other words, an "eponymy principle" stressing competition in Merton's ethos of science, or CUDOS, consisting of communality, universalism, disinterestedness, organizational skepticism, and competition is working to evaluate and reward research achievements produced by the academics as researchers (Merton 1973).

Seventh, institutionalization of academic rankings in the USA functioned as a starting point leading to the internationalization of academic rankings which was initiated in 2003 when agencies such as THES, Shanghai Jiao Tong University, and World News and Report engaged in an academic ranking enterprise with worldwide scope. We, therefore, predict the beginning of international competition among the systems and the institutions. In fact, many countries, including Japan, have become involved in the "rat race": In Japan, the government initiated the twenty-first century COE program in 2002 and the Global COE program in 2008. Similarly in South Korea, there is the World Class University Program, and in China, the 985 project in 1998 (MEXT 2006, 2007; Altbach and Umakoshi 2004; Arimoto 2010a).

12.3.2 Ranking in the World

As has been discussed already, a prototype of worldwide academic ranking emerged about 100 years after it was instituted in the USA. It is important to examine several issues related to this trend, since it is likely to lead to significant differences from what preceded it.

First, the new trend is evident in that the academic community is more and more involved in an international market mechanism. The reputation of institutions as gained in the marketplace has always functioned partly in the realm of the US academic marketplace, but now it is broadly functioning across the world. As a result, ranking has increased its visibility and impact.

Second, the importance of research productivity has increasingly developed together with the internationalization of a "research paradigm." Various kinds of indicators have been used to assess research productivity in addition to those previously used. These include awards such as the Nobel Prize, international awards like the Field Medal, the Albert Lasker Medical Research Award, the Max-Planck-Medaille, the Louisa Gross Horwitz Prize, and John Bates Clark Medal; SCI (Science Citation Index); Eponymous titles as in the Doppler effect, Newton's law, Boil's law, and Mendel's law; scholarships like the Heisenberg Plan, and the post-doctoral program.

As a result, a separation between research and teaching has developed rather than an integration between them, which it is functioning to reduce academic productivity as a linkage between research productivity and teaching productivity. In this context, how to promote such integration and linkage has become a problem to be resolved as soon as possible (Arimoto 2006a, b, 2010a).

Third, the expansion of the academic community beyond internationalization caused by globalization has substantially accelerated a unification of national academic marketplaces. Ranking, which has a close relationship with the shift of COE, creates a huge pyramid of the worldwide academic community beyond national borders. Of course, it is true that the pyramid consists of a variety of components, since the shift of COE occurs according to its deferring levels such as system (country), institution, organization, individual academic, and discipline.

Fourth, related to the first viewpoint, increasing the visibility of the COE brings about a quantitative clarification so that competition intensifies among institutions existing on the status of their COE until the high possibility of scrap-and-build in the process of forming COE is recognized. This kind of trend intensifies competition among not only institutions but also countries so that the emerging countries are forced to participate under the same conditions as advanced countries. Evidence for this is seen in the fact that South Korea and China, both of which were thought to be peripheral to the COE for a long time, are now catching up with more advanced countries in terms of research productivity.

For example, the frequency at which a paper is cited is indicated by the "relative citation impact" (RCI) on the basis of the source. This is compiled by MEXT based on the Thomson Scientific, National Science Indicators, 1981–2006 (Standard version). By country, the RCI is as follows: the USA (1.47), the UK (1.36), Germany (1.21), Canada (1.19), France (1.09), Japan (0.90), South Korea (0.67), China (0.58), India (0.50), and Russia (0.47). Japan's impact has remained low (under 1.0), while those of the UK, Germany, and France have steadily moved upward, approaching that of the USA. Moreover, South Korea and China together with India and Russia have also been basically showing upward trends (MEXT 2008: 55).

Fifth, a market mechanism will intrude on the academic community in accordance with an increasingly intense recruitment of scientists, researchers, and academics with high visibility and reputation, by the ambitious institutions seeking to form the centers of learning. In this process, the growth of both "brain gain" and "brain drain" is internationally recognizable. The separation between research and teaching previously mentioned has progressed so that far more attention is paid to talented academics with high research productivity than to those with high teaching productivity.

Sixth, it has become an increasing problem that the reliability of academic ranking is being questioned (Kobayashi et al. 2005). We fear some indicators used for the evaluation of research productivity are oriented toward the West due to the fact that advanced countries such as the USA and the UK usually occupy higher rankings.

Seventh, the effects of ranking as a result of seeking to achieve COEs have moved from Germany to other countries including Japan via the USA. It is undeniable that the effects are both positive and negative. There are many examples of positive effects: acceleration of research orientation in academia; development of indicators to be used for research productivity; development of evaluation methodologies for research productivity; increase in research productivity in institutions as well as by academics; stimulus for national policy to raise research productivity by way of programs such as the twenty-first century COE program, the Top 30 program, and the World Class University program; social development through scientific development; scrap-and-build in the social stratification of institutions; academic drift and mobility among institutions; and reduction of inbreeding and academic nepotism.

On the other hand, as examples of negative effects we can point to such things as: differentiation of society between "haves" and "have-nots"; separation between research universities and non-research universities; the increase of the research paradigm and the decline of a teaching orientation; the acceleration of application market mechanisms and academic capitalism; a harmful influence of over competition on academics; and an increase of the social pathology of deviant scientific behaviors such as forgery, plagiarism, and fraud.

12.4 Impacts of University Ranking on Academics

12.4.1 Impacts on the Academic Professions

The emergency of academic ranking at an international level in the early twenty-first century has had many effects on various aspects of the academic world including systems, institutions, organizations, and the academics themselves. For example, we can point out that the London Times' ranking published in 2009 caused the following severe shocks.

Examining academic ranking by country, we are able to see both the USA and the UK are ranked at the top of the hierarchy as expected. This is directly related to the history of these two countries in having formed centers of learning for a long time.

Examining academic ranking by region, we can observe that the centers of learning are monopolized by the West, followed by other regions such as Asia, Latin America, and Africa, though these other regions are considered to be peripheral. For example, among the top 100 institutions, the share of the English speaking region, such as the USA, the UK, and Canada, amounts to as much as 67%. Among all institutions, 39 are from Europe, 16 are from Asia, of which the Chinese speaking region, such as Taiwan, China, and Hong Kong, has six institutions (London Times 2009). These data show that the centers of learning formed by advanced countries in the West over a long period are still reflected well in the recent ranking structure.

Further, it is clear that the social stratification is observable at each level of world, region, system, and institution. At the regional level, for example, leading countries are recognizable: the USA and the UK in the West; Japan in Asia, etc. This implies that centers of learning are formed in every region.

In the case of Japan, for example, it is well represented, although it does not belong to the top level of ranking. Thirty Japanese institutions are ranked within the top 600, which corresponds to only 4% of the 770 universities and colleges in Japan. Yet, we would expect 40 if we accept that the ratio of research universities should be 5% (MEXT 2009). As far as the distribution by sector is concerned, 15 of those listed are in the national sector (17.4%), 8 in the private sector (1.3%), and 3 in the public sector (3.5%). The fact that only six institutions, including the University of Tokyo, Kyoto University, Osaka University, Tokyo Institute of Technology, Nagoya University, and Tohoku University, all of which belong to the national sector, are ranked in the top 100, leaves much to be desired.

12.4.2 Academic Productivity by Country

Analyzing academic productivity by country, we can recognize the following noteworthy traits.

12.4.2.1 Rapid Progress of an Academic Drift over 15 Years

Making a comparison of academic productivity in terms of publication by country in the CAP survey of 2007, we can get a ranking with regard to total research productivity (Table 12.2). The top ten consists of South Korea, Japan, Italy, China, Germany, Hong Kong, Norway, Argentina, Canada, and Malaysia. Various items are used in assessing research productivity: publication of book, edited book, article, report, newspaper, patent, computer software, artistic creation, film, and others. Taking the book and the article as examples, we can observe that they have different

Table 12.2 Average productivity in every item by country

	Book	Edited book	Article	Report	Paper	Newspaper	Patent	Computer programs	Artistic	Films	Others	Total
Argentina	0.58 (7)	0.36 (6)	4.45 (14)	2.22 (3)	6.76 (7)	1.52 (6)	0.03 (17)	0.11 (7)	0.40 (5)	0.06 (16)		22.2 (8)
Australia	0.28 (17)	0.19 (18)	6.89 (7)	1.40 (10)	5.79 (9)	1.06 (12)	0.08 (10)	0.08 (14)	0.36 (7)	0.08 (13)	0.18 (15)	18.3 (12)
Brazil	0.55 (8)	0.27 (13)	5.54 (13)	1.53 (7)	5.50 (10)	1.66 (3)	0.04 (15)	0.08 (17)	0.30 (10)	0.16 (6)	0.93 (1)	19.5 (11)
Canada	0.35 (16)	0.28 (12)	6.21 (8)	1.42 (9)	8.16 (1)	1.33 (8)	0.08 (9)	0.11 (6)	0.23 (12)	0.09 (12)	0.62 (2)	20.1 (9)
China	0.83 (4)	0.79 (1)	8.56 (4)	1.32 (11)	2.60 (18)	0.86 (15)	0.30 (4)	0.32 (1)	0.14 (17)	0.07 (15)	0.06 (16)	26.7 (4)
Finland	0.41 (15)	0.36 (9)	5.36 (12)	1.22 (12)	4.67 (15)	1.38 (7)	0.07.(13)	0.10 (11)	0.31 (9)	0.06 (17)	0.35 (11)	17.9 (15)
Germany	0.41 (12)	0.48 (5)	8.76 (3)	2.23 (2)	7.00 (6)	1.62 (5)	0.33 (2)	0.12 (5)	0.41 (4)	0.19 (1)	0.39 (8)	26.4 (5)
Hong Kong	0.48 (10)	0.45 (7)	9.56 (2)	1.64 (6)	7.71 (2)	2.16 (1)	0.20 (5)	0.09 (12)	0.22 (13)	0.12 (9)	0.34 (12)	26.2 (6)
Italy	0.94 (3)	0.48 (4)	8.56 (5)	1.66 (5)	7.52 (3)	1.72 (2)	0.11 (8)	0.09 (13)	0.08 (18)	0.10 (11)	0.26 (14)	29.7 (3)
Japan	1.47 (1)	0.45 (6)	8.54 (6)	1.03 (15)	4.81 (14)	0.92 (13)	0.30 (3)	0.05 (18)	0.68 (2)	0.07 (14)	0.00 (17)	31.3 (2)
South Korea	1.03 (2)	0.65 (2)	10.16 (1)	2.63 (1)	7.15 (5)	1.09 (11)	0.61 (1)	0.10 (8)	0.33 (8)	0.05 (18)	0.50 (5)	36.0 (1)
Malaysia	0.60 (6)	0.33 (10)	4.15 (16)	1.45 (8)	5.95 (8)	0.83 (16)	0.15 (6)	0.10 (9)	0.18 (16)	0.17 (4)	0.38 (9)	19.6 (10)
Mexico	0.41 (13)	0.20 (16)	2.36 (18)	0.56 (18)	3.19 (17)	1.19 (10)	0.04 (16)	0.17 (3)	0.51 (3)	0.17 (2)	0.49 (6)	11.5 (18)
Norway	0.55 (9)	0.26 (14)	5.74 (11)	0.73 (17)	4.91 (13)	1.63 (4)	0.07 (12)	0.09 (15)	0.29 (11)	0.14 (8)	0.58 (4)	18.0 (14)
Portugal	0.65 (5)	0.54 (3)	5.74 (10)	1.78 (4)	7.44 (4)	1.29 (9)	0.12 (7)	0.21 (2)	0.38 (6)	0.17 (3)	0.40 (7)	24.6 (7)
South Africa	0.46 (11)	0.20 (17)	2.84 (17)	0.78 (16)	3.43 (16)	0.62 (18)	0.03 (18)	0.09 (16)	0.22 (14)	0.16 (7)	0.27 (19)	12.9 (17)
UK	0.40 (14)	0.32 (11)	6.05 (9)	1.10 (14)	5.45 (11)	0.78 (17)	0.06 (14)	0.15 (4)	0.20 (15)	0.12 (10)	0.38 (10)	18.0 (13)
USA	0.24 (18)	0.21 (15)	4.26 (15)	1.11 (13)	5.41 (12)	0.89 (14)	0.08 (11)	0.10 (10)	1.23 (1)	0.16 (5)	0.59 (3)	14.6 (16)
Total	0.61	0.21	4.26	1.11	5.41	0.89	0.08	0.1	1.23	0.16	0.59	14.6

Source: Daizen 2010)

Notes: Average productivity is calculated by weighting different academic outputs and aggregating the scores: 10 points for each book; 5 points for edited book; 1 point for each book chapter or journal article; 3 points for each research report; 0.5 points for each paper presented to an academic association, patent, computer, artistic activity, or film; 0.3 points for each newspaper article; others are not included in the total of average productivity

ranking orders: For the book, the order of the top five is Japan, South Korea, Italy, China, and Norway, while for the article, the order is South Korea, Hong Kong, Germany, China, Italy, and Japan. In comparison, in the 1992 survey, the top ten consisted of Japan, the Netherlands, Sweden, Germany, Chile, Israel, the USA, the UK, Brazil, and Australia (Arimoto and Ehara 1996: 172).

In the overall ranking, South Korea has achieved a major breakthrough moving from 11th in 1992 to the top in 2007. The reason for this successful outcome is probably due to a series of national projects such as the first BK21(1999–2005) and the second BK21(2006–2012). As far as the outcome of the former is concerned, quantitative and qualitative enhancement of research is shown in the fact that the number of articles counted by SCI (Science Citation Index) doubled from 3,765 in 1998 to 7,281 in 2005, and in the same period, national ranking of SCI up from 18th to 12th. An impact factor of SCI in the field of science and technology increased from 1.9 in 1999 to 2.43 in 2005. Taking into account the WCU (World Class University) project (2008–2012), which was introduced in 2008 after the CAP survey was conducted in 2007, we expect to see an even more successful outcome in the future (Umakoshi 2010: 75–95).

12.4.2.2 Japan is Keeping a High Ranking

The huge increase in South Korea's ranking has displaced Japan to the second position. While productivity in Japan may well have been affected by a series of structural changes (see below), the disciplinary bias in sampling, which over-represents the medical sciences, may also be a contributory factor. In other words, the sampled share of each disciplinary area is as follows: Humanities (13.5%), Social sciences (13.6%), Natural sciences (18.8%), Engineering (24.5%), Medical sciences (22.7%), and Teacher training (7.0%). It follows that sampling is the second highest in the medical sciences following engineering in Japan, and in an international comparison of share, Japan is the highest in this disciplinary composition when compared with Australia (19.3%), Brazil (18.5%), Norway (17.8%), Germany (15.6%), etc. As Table 12.3 shows, Japan has the highest productivity in the medical sciences with a score of 51.5 so that both high percentage and productivity in this field seem to be significant when we note Japan is keeping its high ranking.

12.4.2.3 Emergence of Italy, China, Norway, etc. as New Faces

Of the 18 countries participating in the 2007 survey, Italy has achieved a high rank. Italy, China, and Norway did not participate in the survey in 1992. Germany's research productivity provides a remarkable contrast with that of the USA and the UK, although these three countries have been highly ranked COE members for a long time. Germany established the original COE in the nineteenth century and since then has maintained a high position until today. It is interesting to note that Germany had a high ranking productivity in both the 1992 and 2007 surveys,

Table 12.3 Average research productivity by discipline and country

	Humanities	Social sciences	Natural sciences	Engineering	Medical sciences	Teacher training	Total
South Korea	32.2 (1)	34.6 (1)	35.3 (1)	41.3 (1)	47.6 (2)	31.4 (2)	36.0 (1)
Japan	21.8 (9)	24.9 (5)	26.6 (4)	26.5 (5)	51.5 (1)	26.5 (3)	31.3 (2)
Italy	30.5 (4)	29.3 (2)	26.1 (5)	30.9 (3)	39.3 (4)	36.6 (1)	29.7 (3)
China	25.7 (6)	28.3 (3)	25.1 (8)	28.4 (4)	27.0 (7)	22.9 (7)	26.7 (4)
Germany	30.8 (3)	27.5 (4)	26.7 (3)	20.1 (10)	28.8 (6)	28.8 (5)	26.4 (5)
Hong Kong	20.2 (11)	18.9 (11)	31.2 (2)	39.3 (2)	38.2 (5)	24.4 (4)	26.2 (6)
Portugal	26.7 (5)	21.8 (6)	25.7 (7)	23.9 (8)	39.6 (3)	15.5 (13)	24.6 (7)
Argentina	32.0 (2)	20.3 (7)	19.6 (11)	17.0 (14)	22.5 (10)	22.9 (6)	22.2 (8)
Canada	17.1 (16)	18.3 (13)	21.0 (10)	24.8 (7)	23.8 (8)	21.3 (9)	20.1 (9)
Malaysia	23.1 (8)	19.8 (8)	25.9 (6)	15.0 (16)	15.6 (16)	21.7 (8)	19.6 (10)
Brazil	23.7 (7)	17.7 (15)	21.2 (9)	19.4 (12)	19.0 (12)	16.7 (12)	19.5 (11)
Australia	18.6 (12)	19.6 (9)	17.4 (13)	20.0 (11)	17.8 (13)	16.8 (11)	18.3 (12)
UK	17.4 (14)	18.5 (12)	19.4 (12)	20.9 (9)	15.7 (15)	11.1 (18)	18.0 (13)
Norway	20.3 (10)	19.4 (10)	14.6 (16)	16.9 (15)	20.1 (11)	18.9 (10)	18.0 (14)
Finland	17.9 (13)	18.1 (14)	16.0 (15)	17.3 (13)	23.2 (9)	14.1 (14)	17.9 (15)
USA	13.0 (17)	12.2 (18)	17.2 (14)	25.1 (6)	16.9 (14)	11.6 (17)	14.6 (16)
South Africa	12.2 (18)	16.4 (16)	11.6 (18)	7.0 (18)	5.4 (18)	12.5 (15)	12.9 (17)
Mexico	17.2 (15)	13.1 (17)	12.5 (17)	7.1 (17)	11.0 (17)	11.6 (16)	11.5 (18)
Total	22	20.8	22.3	22.7	26.6	19.2	22.2

while in the London Times ranking, only four German institutions are ranked within the top 100, far fewer than their counterparts in the USA and the UK (London Times 2009).

On the other hand, in the CAP survey of research productivity, both the USA and the UK are ranked far lower than expected. In analyzing the results of the 2007 survey, we find that Germany is ranked within the top five, while both the USA and the UK are ranked 16th and 13th, respectively. In the individual categories for books published, the rankings are 18th and 14th, respectively, and for articles 9th and 15th, respectively (Table 12.2). These two countries therefore revealed surprisingly low research productivity in the 2007 survey, which was based on academics' responses to questions about research productivity rather than on an external review undertaken in the London Times survey. Incidentally, in the 1992 survey, the UK was at 8th and the USA was at 7th position (Arimoto and Ehara 1996). Why is their productivity low? Probably, it is a reflection of the small sampling of academics from the group of research universities in the CAP survey.

12.4.2.4 Academics' Increasing Commitment to a Research Orientation

In the 1992 survey, three groups were distinguished with regard to academics' orientation toward research and teaching: a German type, an Anglo Saxon type, and a Latin type (Arimoto and Ehara 1996). The German type consists of countries such as Germany, the Netherlands, Sweden, South Korea, and Japan, characterized by a strong research orientation. The Anglo Saxon type consists of countries (and a region) such as the USA, the UK, Australia, Hong Kong, and Taiwan, showing almost equal orientation to both research and teaching. The Latin American type consists of countries such as Brazil, Argentina, Chile, and Russia, characterized by a strong teaching orientation.

The recent data (2007) show that in both the Anglo Saxon type and the Latin American type, there is an increased orientation to research, with a diminished teaching orientation (Arimoto 2010a). This trend means in part that there will be a large growth in the German type within 15 years and indicates the increasing effect of international academic ranking on the much greater involvement by academics in research orientation throughout the world. It follows that this trend obviously implies a converse effect on the integration of research and teaching.

12.4.3 Academic Productivity by Structural Factors

Even if the research productivity of the Japanese academic is still high according to the quantitative data shown above, it may still be declining. In other words, due to a series of changes in higher education policies undertaken continuously over the past 15 years, which has brought about a separation between research orientation and teaching orientation, the research orientation in Japan appears to be gradually

weakening. This trend is contrary to the Humboldtian ideal of integration of research and teaching (Von Humboldt 1910; Boyer 1990; Clark 1997; Ushiogi 2008; Arimoto 2010a). As a result, it has caused a type of anomie in the realm of academics' consciousness, in the sense that there is a widening discrepancy between research and teaching rather than an integration between them.

For example, there have been a series of policies to reinforce this situation as follows (UC 1998; CEC 2005; Arimoto 2007, 2010b, c):

- A system-level measure by way of the Science and Engineering Basic Law (1995) and Planning (1996).
- Higher education policies with a focus on the COE program, the Top 30 program, and a teaching-oriented type of Faculty Development (FD). By introducing these policies, a differentiated society appeared in the academic community with a widening discrepancy between research universities and non-research universities.

Parallel to this trend, academic productivity, consisting of both research productivity and teaching productivity, has declined considerably. The effect of this disintegration of research and teaching, instead of an integration between them, will become more evident in the future, because academic ranking is based on research orientation rather than teaching orientation as is shown in the trends of many of the countries participating in the CAP survey (Arimoto 2010a). An extension of this trend in the future is likely to cause further segmentation both among institutions and among academics.

The reaction of the academic profession is worthwhile noting because academics belong to a range of social groupings such as those identified as system, sector, section, tier, hierarchy, status, age group, or gender which must define their specific responses. That the results of the London Times and the CAP survey do not converge is natural when we consider the many differences in academics' reactions. Both surveys deal with academic productivity, especially research productivity. In the London Times, research productivity is defined by an external evaluation using various indicators, while research productivity in the CAP is defined by an internal evaluation on the basis of the consciousness of the academics who respond to the questionnaire. Such consciousness is, as it were, a collaboration of the various factors of system, institution, sector, section, etc. to which the individual academic belongs. These factors are each examined below.

12.4.3.1 System

The number of higher education systems may well equate to the number of countries. Some of them are in advanced countries and some are in developing countries. It is assumed that advanced countries are likely to be positive toward the increasing productivity rate because the concept of the COE is familiar to these systems on the basis of either having the centers of learning, or of seeking to become centers of learning in the future. Conversely for the developing countries, it is less likely to be

positive because the concept is thought to be less familiar to their systems. Those systems that already have centers of learning are likely to realize high academic productivity, but it is interesting, as the CAP survey reveals, that the centers of learning in the USA, and the UK are not necessarily ranked highly in the CAP, possibly because of biased sampling.

12.4.3.2 Sector

In many countries, the higher education system is divided into two or more sectors. In Japan, the system is divided into three sectors: the national, the public, and the private sector. The national sector is accorded the highest prestige, followed by the public sector and the private sector. In the USA., there are two categories: the state and the private sector. In general, the private sector is considered to be more prestigious than the public. In Germany, the state sector prevails almost exclusively with several states sharing a largely equivalent prestige (Arimoto 1996). In the UK, the public sector has the highest prestige, with Oxbridge at the top, and this structure is similar to that of France and Japan (Clark 1983). In South Korea, there are two sectors with Seoul University at the top, and in China, multiple sectors form a hierarchy headed up by Beijing University and Tsinghua University.

In the London Times survey, we can recognize the correlation between the high ranking institutions and their ranking in every country. In the case of Japan, for example, all the institutions ranked within the top 100 are in the category of research universities in the national sector. To become a high ranking institution, a research university clearly needs to be connected to the national sector.

12.4.3.3 Section

The category referred to as section relates to the academic discipline, since the section consists of faculty, department, and chair on the basis of discipline. Disciplines in the field of the natural sciences tend to have a well-developed scientific codification so that clear standards are used to assess the quality of academic productivity. Physics provides a typical applied example of this (Zuckerman and Merton 1971). By contrast, the humanities and social sciences tend to have a less developed scientific codification so that the quality of academic productivity is frequently assessed by ambiguous standards. As a result, it is not surprising that the natural sciences are apt to have an international orientation, while the humanities and social sciences will have more of a local orientation. For example, articles in the natural sciences are usually written in English – a language common to the international academic community. Conversely, an article written in Japanese is unlikely to be read by researchers or scientists in the international academic community.

The humanities and social sciences are more likely to be committed to the cultures and traditions particular to their individual countries, or local regions. Academics in these disciplines often write articles in the languages familiar to those

cultures and traditions. In Japan, there are many academic journals serving these disciplines which are published by universities and colleges. Many academics contribute articles in Japanese. These articles receive very little attention by foreign scholars who do not read Japanese, even if these articles are of a high quality, comparable to that of a COE in line with international standards. Accordingly, international recognizable ranking is more easily attained by Japanese academics in the field of natural sciences compared to their counterparts in humanities and social sciences (Arimoto 1994).

It may be said that the value of "universalism" is working in the natural sciences, while "particularism" is working in the humanities and social sciences. This appreciable difference in impact may explain the high productivity of Japanese academics in the natural sciences, and especially in medical sciences. As has already been discussed, over-sampling of academics in this section in addition to their outstanding productivity appears to have been a major factor in raising the average productivity of all Japanese academics to that of second best ranking according to the CAP survey.

12.4.3.4 Tier

Tier indicates the level of the academic and intellectual program and has a close relation to the stratification of knowledge. The level of difficulty of content in a discipline is described as beginning, intermediate, or advanced. Such content difficulty is aligned with curricula in the schools as well as the universities and colleges whether it is elementary school, middle school, senior high school, or university and college. In the case of universities and colleges, the content difficulty of the curricula is higher in the postgraduate level than that in the undergraduate level, because the former has a closer connection with advanced research than the latter. Therefore, we assume that academics in the postgraduate tier are likely to be very conscious of ranking, and in particular those in the research universities. They are not only conscious of the ranking system but also know the exact global locations of the COE.

12.4.3.5 Hierarchy

Hierarchy is a social stratification incorporating the academics, institutions, and systems that form around a discipline and extend from its COE to the periphery. As discussed above, the hierarchical structure is likely to be unitary in the natural sciences, but likely to be pluralistic in the humanities and social sciences. In terms of academic productivity, an institution can become the COE if it attracts many prestigious researchers in a specific discipline. The earlier description of the Departments of Sociology in the USA is one example of this. In similar systems such as those of the USA and the UK, many institutions are capable of becoming COE.

Such COEs can be identified by various indicators, and ranking is actually a result of the application of these indicators. It is no exaggeration to say that ranking occurs in a hierarchy, in which usage of indicators substantially changes the given ranking as well as the given hierarchy. Currently ranking is based on the structure that weights research productivity higher than teaching productivity. The same ranking would not be seen if the weight were predominantly on teaching productivity rather than research productivity.

In a recent (2010) conversation with a professor at the University of Paris in France, he suggested that the reason why only two French institutions are in the top 100 of the London Times ranking is that they operate according to different values from those that influence the ranking. Historically, French universities focused on teaching, while the academy focused on research (Clark 1983; Arimoto 1996). The USSR and, at an early stage, China imported the French system in which the weight on teaching differed from that in American and British universities where they attach special importance to research and teaching. Hence, one can understand why French universities may pursue structurally different academic productivity from that of the Anglo Saxon environment.

12.4.3.6 Hierarchy of Position

There is also a hierarchy of position which relates to academics' social stratification. A professor's prestige and academic productivity are thought to be high whenever promotion is based on competition. Accordingly, professors are most susceptible to ranking during their academic careers, knowing the significance of the ranking. In this climate, junior academics, who are engaged in competition with senior academics, seek to enhance their own academic productivity.

Academics aspiring to upward mobility are clearly responding positively to academic ranking. Academics, particularly full professors, who climb to the highest ranks in the "pyramidal type" of professoriate population in Western universities and colleges, are thought to be sensitive to academic ranking. In contrast, full professors in the "chimney type," or reverse pyramidal type, of Japanese universities and colleges, are thought to be less sensitive. About 40 years ago, in Western universities in countries such as the USA, the UK, Germany, and France, the ratio of professorial to junior positions reflected the pyramidal type and this is still the case today. As Morikazu Ushiogi has pointed out, in Japan it was a chimney type, categorized by Michiya Shinbori 40 years ago, but it has changed to the reverse pyramidal type today (Shinbori 1965; Ushiogi 2009).

12.4.3.7 Research Money from Outside Academia

Research money is one of the most important factors influencing the academic profession's reaction to ranking. It is difficult to improve academic productivity without research money. One reason why academics in the USA are competitive in

research productivity may well lie in the culture, climate, and atmosphere of its society generally as well as in academia. The relationship between society and academia is based on a market mechanism operating at a level in the USA, well ahead of that in any other countries.

The connection between academia and research money from outside is strong. Research money from outside is the highest in the USA (0.380), followed by South Korea (0.337), and Japan (0.308), while in the UK (0.232) and Germany (0.141) it is lower (Table 12.4). It is perhaps surprising to find that German academics have high productivity with less outside money than the total average (0.165), while South Korean and Japanese academics have high productivity with considerable outside money. However, it is even more surprising that American academics have lower productivity despite the fact that they receive the highest level of outside money.

12.4.4 Academic Productivity by Faculty Factors

12.4.4.1 Ph.D. Degree

In the academic environment, the Ph.D. degree has become the "union card" as described earlier, and it is also related to high research productivity. The degree is seen as valuable for recruitment and promotion to higher ranks but what about its effect on productivity? Table 12.4 shows that in all countries, Ph.D. holders are more productive than other degree holders. This trend is typical of Japan (36.0%).

12.4.4.2 Age

Age is generally associated with a hierarchy of position. In general, junior academics occupy positions in the lower levels of the hierarchy, while the senior academics occupy higher positions although at times there are exceptions to this. In a system in which both promotion and upward mobility are frequently related to competition and selection processes, academics are forced to take academic ranking into account. On the other hand, in a system not driven by these mechanisms, the academics can be less influenced by ranking.

In the pyramidal structure, the situation is competitive; this is far less so in the chimney type and the reverse pyramidal structure. The cluster of research universities in the U.S. is highly competitive in the promotion process since it corresponds to the pyramidal type, while the counterpart in Japan is less competitive since it retains the characteristics of a reverse pyramid type. American academics are inclined to respond far more positively to academic ranking than Japanese academics. Academics over the age of 40 years are more productive than those under 40 years in the five countries listed in Table 12.4. As far as these data are concerned,

Table 12.4 Characteristic of total productivity by country

Country	Productivity (total)	Gender M.	Gender F.	Over 40	Under 40	Degree Dr.	Degree (others)	Research money
South Korea	35.3 (1)	34.1 (1)	40.2 (1)	36.2 (1)	32.4 (1)	28.8 (4)	20.5 (4)	0.337 (2)
Germany	26.7 (3)	28.6 (3)	19.4 (8)	33.6 (3)	14.0 (14)	31.2 (2)	8.0 (17)	0.141 (12)
Japan	26.6 (4)	26.6 (7)	24.0 (5)	27.8 (7)	17.9 (7)	36.0 (1)	20.1 (5)	0.308 (3)
UK	19.4 (12)	20.8 (11)	16.2 (14)	20.9 (13)	14.7 (13)	17.9 (15)	4.4 (18)	0.232 (8)
USA	17.2 (14)	17.3 (15)	17.2 (11)	17.7 (16)	15.1 (12)	24.7 (10)	17.5 (7)	0.380 (1)
Total (18 countries)	22.3	23.7	19	25	16.6	24.7	17.5	0.165

Japanese academics are productive in both generations, despite the existence of the reverse pyramidal model.

12.4.4.3 Gender

The significance of a gender in relation to ranking lies in the fact that inequality exists between men and women in the academic profession. Japanese female academics are proportionately fewer in the total population of academics than their counterparts in other countries. It follows that they are disadvantaged in the processes of establishing their status. By contrast, American female academics have a greater advantage in the recognition of their status.

Once they have obtained a higher status, female academics are able to take on many more opportunities to participate in research productivity at the international level. As shown in Table 12.4, female academics' productivity (19.0) is lower than the male academics' productivity (23.7), when data from 18 countries are compared. This is also true for countries at an individual level. For example, in Japan, female productivity is 24.0 where as the male productivity is 26.7. A similar trend is seen in Germany, the UK, and South Korea, however, is exceptional in that female productivity at 40.2 is higher than male productivity at 34.1. In the USA, both genders show almost the same productivity. It follows that female academics in South Korea and the USA are making positive contributions to academic ranking.

The situation of female academics in the USA has changed dramatically. They had a significantly lower status until 50 years ago. The ratio of female academics employed at Harvard University in 1959–1960 remained low at all professional levels and especially at the position of full professor (Keller and Keller 2001: 278) (Table 12.5). Affirmative action was introduced in 1976 and the ratio still remained low though no legal control had existed before that time of 1959–1960. It is clear that up to that time, opportunity for female academics was extremely limited even in one of the most progressive institutions in the USA.

As has been discussed, the attitude and reactions of the academic profession toward academic ranking are not simple to establish, because they are affected by many factors. Taking this into account, we can expect to approach this theme by making a combination of these factors in the next step of research in the future.

Table 12.5 The ratio of female academics employed at Harvard in 1959–1960

Position	Total	Women
Professor	427	4 (0.9%)
Associate professor	118	8 (6.8%)
Assistant professor	299	8 (2.7%)
Instructor	529	52 (9.8%)
Lecturer	196	28 (14.3%)
Research staff	769	107 (14.7%)
Teaching fellows	597	76 (12.7%)
Clinical (HMs)	236	8 (3.4%)

12.5 Concluding Remarks

This chapter has dealt with many problems related to the main theme, by analyzing the following issues and drawing conclusions.

1. Academics are engaged in functions of knowledge, especially research—as the discovery of knowledge and teaching-as the dissemination of knowledge; and their outputs of research and teaching productivity. Academic productivity is an integration of these two functions, and is the most important role the academic profession is expected to pursue in the age of transformation from A1 to A2.

   By paying attention to academic productivity, we can realize the formation of COEs and their condition, structure, and function. A center of learning was built on the teaching and productivity reputation of institutions and academics in the medieval universities. In modern universities, by comparison, it has been built on research productivity. This seems to suggest that a research paradigm has prevailed in the academic community since the institutionalization of German universities in the nineteenth century.
2. American universities established graduate schools in the nineteenth century at the A1 stage when they imported the German model. They enhanced their research productivity to become the major COEs in the world. In order to be able to invite distinguished scholars to their staff, they controlled inbreeding and academic nepotism as much as possible. At the same time, they attempted a series of academic reforms in order to enhance research and raise research productivity.

   On the other hand, many countries failed to introduce similar reforms. Japanese universities paid a great deal of attention to German universities, inviting prominent scholars to visit and also sending students abroad. However, they were not successful in introducing reforms leading to high research productivity, having encouraged inbreeding rather than controlling it for many years.
3. In the USA, universities as well as academics responded positively to the institutionalization of academic ranking. The reason lay mostly in their competitiveness which sought to catch up with and exceed Germany's high level of research productivity. There was considerable competition among departments for a primacy ranking. This trend introduced ranking into the academic marketplace for the first time, stimulating academic drift between institutions and a scrap-and-build attitude toward institutional ranking.

   Based on this trend, the research universities strengthened their positions in the academic marketplace and realized higher positions in the hierarchy of higher education institutions. Accordingly, the academic ranking introduced in the USA in the A1 era was a foretaste of what was to come with the start of the A2 era in the twenty-first century.
4. Emergence of worldwide academic ranking is establishing a hierarchy of higher education institutions as a unified pyramidal structure around the world, in which the West-centered structure focused on the USA and the UK is prevailing. As a result, some positive and negative effects have occurred in the academic community.

5. There is a difference between objective and subjective evaluation in academic ranking. For example, the USA and the UK, which are ranked highly by the London Times and other surveys, are ranked lower when compared against the 18 countries of the CAP survey.

   An objective evaluation usually emphasizes research rather than teaching, and research productivity rather than teaching productivity. Both research and teaching are indispensable in the academic community. Nevertheless, the fact that a teaching orientation has decreased and a research orientation has increased in the CAP survey suggests a close relationship to the emerging worldwide academic ranking.

   An emerging academic ranking is affecting in manifest and latent function level the consciousness of the academic profession and also national higher education policy. The academics in South Korea and some other countries have responded positively to this kind of trend over the past 15 years. In Japan, the quality of academic productivity is slowing down, even though academics have constantly maintained a high level of productivity over the past 15 years. The reason is attributable to the effects of conflicts between the academics' traditional research orientation and the national government's higher education policy, demanding all academics to improve their teaching orientation.
6. The reactions of the academic profession to academic ranking are caused by various factors: system, sector, section, tier, hierarchy, hierarchy of position, age, and gender. Consequently, the reactions of the academic profession to academic ranking are complicated. It is said that we can observe the problems, which objective evaluation can hardly understand, through subjective evaluation. However, more detailed observation of the reactions of the academic profession in regard to academic ranking is needed to provide a greater in-depth understanding.

References

Altbach, P. G., & Umakoshi, T. (Eds.). (2004). *Asian universities: Historical perspectives and contemporary challenges*. Baltimore: The Johns Hopkins University Press.

Amano, I. (1984). Method of university classification (in Japanese). In T. Keii (Ed.), *Study of university evaluation*. Tokyo: University of Tokyo Press.

Arimoto, A. (1981). *Sociology of academics (in Japanese)*. Tokyo: Gakubunsha Publishing Co.

Arimoto, A. (1987). *Study of Merton sociology of science: Formation and development of its paradigm (in Japanese)*. Tokyo: Fukumura Publishing Co.

Arimoto, A. (1994). *Academic discipline and academic productivity: A case study based on a survey of department chairpersons in research university in the U.S.A., Research in Higher Education* (Vol. 23, pp. 27–47). Hiroshima: RIHE, Hiroshima University.

Arimoto, A. (Ed.). (1996). *Study of center of learning (in Japanese)*. Tokyo: Toshindo Publishing.

Arimoto, A. (2005). *Academic profession and FD in the U.S. and Japan*. Tokyo: Toshindo Publishing.

Arimoto, A. (2006a). *Institutionalization of faculty development with a focus on Japan, reports of changing academic profession project workshop on quality, relevance, and governance in the*

changing academia: International perspectives (COE publication series, Vol. 20, pp. 3–20). Hiroshima: RIHE, Hiroshima University.

Arimoto, A. (2006b). *The changing academic profession in Japan: Its origins, heritage and current situation, reports of changing academic profession project workshop on quality, relevance, and governance in the changing academia: International perspectives* (COE publication series, Vol. 20, pp. 183–194). Hiroshima: RIHE, Hiroshima University.

Arimoto, A. (2007). National research policy and higher education reforms with focus on Japanese case. In Sorlin Sverker & Vessuri Hebe (Eds.), *Knowledge society vs. knowledge economy: Knowledge, power, and politics* (pp. 175–197). New York: Palgrave Macmillan.

Arimoto, A. (Ed.). (2008). *The changing academic profession in Japan (in Japanese) (CAPJ)*. Tokyo: Tamagawa University Press.

Arimoto, A. (2009a, September). *Changing academic profession in the world from 1992 to 2007*. Keynote presented at the CAP seminar, Garden Palace, Hiroshima, February, 2009 (RIHE International Seminar Report No. 13, pp. 1–37). Hiroshima: RIHE, Hiroshima University.

Arimoto, A. (2009b). The competitive environment of academic productivity and the academic research enterprise in the case of Japan. *Asia Pacific Education Review, 10*(1), 29–46.

Arimoto, A. (2010a, January 13–14). *Differentiation and integration of research, teaching and learning in the knowledge society: From the perspective of Japan.* Paper presented to the CAP Hiroshima Conference, Hiroshima.

Arimoto, A. (2010b). Japan. In D. Dill & F. A. Van Vught (Eds.), *National innovation and the academic research enterprise* (pp. 115–147). Baltimore: The Johns Hopkins University.

Arimoto, A. (2010c). The academic profession and the management university: An international comparative study from Japan. *European Review, 18*, 117–139.

Arimoto, A., & Ehara, T. (Eds.). (1996). *International comparison of the academic profession (in Japanese)*. Tokyo: Tamagawa University Press.

Bälz, E. (1979). (Toku Bälz (Ed.). Rytaro Suganuma, Trans.) *Daiary of Bälz (in Japanese)*. Tokyo: Iwanami Publishing Co.

Becher, T. (1981). Towards a definition of disciplinary cultures. *Studies in Higher Education, 6*, 109–122.

Becher, T. (1989). *Academic tribes and territories: Intellectual enquiry and the cultures of the disciplines*. Milton Keynes: Open University Press.

Becher, T., & Parry, S. (2007). The endurance of the disciplines. In I. Bleiklie & M. Henkel (Eds.), *Governing knowledge: A study of continuity and change in higher education-A Fertscrift in Honour of Maurice Kogan* (pp. 133–144). Dordrecht: Springer.

Ben-David, J. (1977). *Centers of learning: Britain, France, Germany, United States: An essay*. New York: McGraw-Hill.

Bleiklie, I., & Henkel, M. (Eds.). (2005). *Governing knowledge: A study of continuity and change in higher education: A Festshrift in Honor of Maurice Kogan*. Dordrecht: Springer.

Boyer, E. L. (1990). *Scholarship reconsidered: Priorities of the professoriate*. Princeton: Carnegie Foundation.

Brubacher, J., & Rudy, W. (1968). *Higher education in transition*. New York: Harper & Row.

Bush, V. (1945). *Science, the endless frontier*. Washington, DC: United States Government Printing Office.

Carnegie Commission on Higher Education. (1976). *A classification of higher education*. Berkeley: Carnegie Foundation. Princeton: Princeton University Press.

CEC (Central Education Council). (2005). *Future vision of higher education in Japan (in Japanese)*. Tokyo: Central Educational Council.

Clark, B. R. (1983). *Higher education system: Academic organization in cross-national perspective*. Berkley: University of California Press.

Clark, B. R. (1987). *The academic life: Small worlds, different worlds*. Princeton: The Carnegie Foundation for the Advancement of Teaching.

Clark, B. R. (1995). *Places of inquiry: Research and advanced education in modern university*. Berkeley: University of California Press.

Clark, B. R. (1997). The modern integration of research activities with teaching and learning. *Journal of Higher Education, 68*(3), 241–255.
Clark, B. R. (1998). *Creating entrepreneurial universities: Organizational pathways of transformation*. Oxford: Pergamon/Elsevier.
Clark, B. R. (2008). *On higher education: Selected writings, 1956–2006*. Baltimore: The Johns Hopkins University Press.
Cole, J. R., & Cole, S. (1973). *Social stratification in science*. Chicago: The University of Chicago Press.
Daizen, T. (2010). Research productivity (in Japanese). In A. Arimoto (Ed.), *Changing academic profession in the world, forthcoming*. Tokyo: Tamagawa University Press.
Drucker, P. (1993). *Post capitalist society*. Oxford: Butterworth-Heinemann.
Geiger, L. (1986). *To advance knowledge: The growth of American research universities since world war II*. New Brunswick: Transaction.
Geiger, L. (2004). *Knowledge and money: Research universities and the paradox of the marketplace*. Stanford: Stanford University Press.
Gibbons, M., Limoges, C., Nowotny, H., Schwartzman, S., Scott, P., & Trow, M. (1994). *The new production of knowledge: The dynamics of science and research in contemporary societies*. London: SAGE Publications.
Graham, H. D., & Diamond, N. (1997). *The rise of American research universities: Elites and challengers in the postwar era*. Baltimore/London: Johns Hopkins University.
Jacob, M., & Hellstrom, T. (Eds.). (2000). *The future of knowledge production in the academy*. Buckingham: The Society for Research into Higher Education/Open University Press.
Jones, L. V., Lindzey, G., & Coggeshall, P. E. (Eds.). (1982). *An assessment of research-doctorate programs in the United States* (Vol. 1–5). Washington, DC: National Academic Press.
Keller, M., & Keller, P. (2001). *Making harvard modern: The rise of America's university*. Oxford/New York: Oxford University Press.
Kobayashi, M., et al. (2005). *Comparison of academic ranking of world universities (in Japanese), Shigaku Koto Kyouiku Kenkyu Sosho* (pp. 1–50). Tokyo: Shigaku Koto Kyoiku Kenkyusho (Research Institute for Private Higher Education).
Kogan, M. K., Bauer, M., Bleiklie, I., & Henkel, M. (Eds.). (2006). *Transforming higher education: A comparative study* (2nd ed.). Dordrecht: Springer.
Light, D. W., Jr. (1974). The structure of the academic professions. *Sociology of Education, 47*, 2–28.
London Times (2009). QS Quacquarelli Symonds. (www.topuniversities.com)
Merton, R. K. (1973). In N. W. Storer (Ed.), *The sociology of science: Theoretical and empirical investigations*. Chicago: University of Chicago Press.
MEXT (2006). *Annual report on the promotion of science and technology,* (FY 2006). Tokyo: MEXT.
MEXT (2007). *White paper on science and technology* (FY 2007). Tokyo: MEXT.
MEXT (2008). *White paper on science and technology 2008*. Tokyo: MEXT.
MEXT (2009). *School basic statistics* (FY 2009). Tokyo: MEXT.
Nakayama, S. (1978). *Birth of imperial university (in Japanese)*. Tokyo: Chuoukoron Publishing Co.
Oleson, A., & Voss, J. (Eds.). (1979). *The organization of knowledge in modern America, 1860–1920*. Baltimore: The Johns Hopkins University Press.
Parry, S. (2007). *Disciplines and doctorates*. Dordrecht: Springer.
Parsons, T., & Platt, G. M. (1973). *The American university*. Massachusetts: Harvard University Press.
Pierson, G. W. (1952). *Yale college: An educational history 1871–1921*. New Haven: Yale University Press.
RIHE (2009). *The changing academic profession over 1992–2007: International, comparative, and quantitative perspective* (RIHE International Seminar Report No. 13). Hiroshima: RIHE, Hiroshima University.
Rudolph, F. (1962). *The American college and university*. New York: Knopf.

Ruffner, J. A. (Ed.). (1977). *Eponyms dictionary index*. Detroit: Gale Research Company.
Shils, E. (1979). The order of leaning in the United States: The ascendancy of the university. In A. Oleson & J. Voss (Eds.), *The organization of knowledge in modern America, 1860–1920*. Baltimore: The Johns Hopkins University Press.
Shinbori, M. (1965). *Academic marketplace in Japan (in Japanese)*. Tokyo: Toyokan Publishing Co.
Shinbori, M. (1973). *Research of academic productivity (in Japanese)* (Daigaku Ronshu, Vol. 1, pp. 11–19). Hiroshima: RIHE, Hiroshima University.
Shinbori, M. (Ed.). (1985). *Evaluation of academic achievement: Eponymy phenomena in science (in Japanese)*. Tokyo: Tamagawa University Press.
Slaughter, S., & Leslie, L. L. (1997). *Academic capitalism: Politics, policies, and the entrepreneurial university*. Baltimore: The Johns Hopkins University Press.
UC (University Council). (1998). *21st university vision and direction for reforms (in Japanese)*. Tokyo: University Council.
Umakoshi, T. (2010). *Dynamism of Korean academic reforms (in Japanese)*. Tokyo: Toshindo Publishing Co.
Ushiogi, M. (1984). *Challenge of Kyoto imperial university (in Japanese)*. Tokyo: Chuoukoron Publishing Co.
Ushiogi, M. (2008). *Is the end of Humboldtian ideal? New dimension of modern University (in Japanese)*. Tokyo: Toshindo Publishing Co.
Ushiogi, M. (2009). *Professor as profession (in Japanese)*. Tokyo: Chuoukoron Publishing Co.
Veysey, L. R. (1965). *The emergence of American universities*. Chicago: The University of Chicago Press.
Von Humboldt, W. (1910). On the spirit and the organizational framework of intellectual institutions in Berlin (Edward Shils, Trans.). *Minelva*, *8* (1970), 242–250.
Yamanoi, A. (Ed.). (2007). *Academic market place in Japan (Japanese)*. Tokyo: Tamagawa University Press.
Zuckerman, H., & Merton, R. K. (1971). Patterns of evaluation in science: Institutionalization, structure and functions of the referee system. *Minerva, 9*(1), 66–100.

Chapter 13
The Future of University Rankings

Ulrich Teichler

13.1 The Amazing 'Popularity' of University Rankings

In analysing the abundant meta-evaluative literature on university rankings, we note an impressive range of arguments and analyses of the concepts, methods, results, perceptions and possible impact of such kinds of activities to put individual institutions somehow 'on a map'. In various areas of research, we often note that a certain approach draws substantial attention because it is viewed as very ambitious and promising and as deserving further enhancement through the involvement of the brightest scholars in the respective field. In the case of university rankings, however, most experts would agree that the great attention paid to this domain by many experts is not an indication of respect for high quality analysis. Rather, rankings draw attention as a consequence of a seemingly paradoxical mixture of conceptual and methodological weakness on the one hand and political power on the other to influence the views of the 'map' of higher education and to elicit activities aimed at changing the existing 'map'.

The expert literature obviously does not spend much time and energy on defining rankings. 'Rankings' exist, and efforts to clarify their definition would be futile, given the lack of precision of what is meant by that term. However, we note that ranking studies are usually described as data presentations with three general features:

- Ranking studies are activities of vertical sorting. Rank lists or scales are established according to 'very good', 'high-quality', 'excellent', 'world class', 'renowned' or whatever the positive end in higher education might be called.
- Ranking studies carry out an inter-institutional comparison. Higher education institutions or their sub-units (departments, etc.) are compared, as a rule, across all higher education institutions within a country, region or worldwide.
- Ranking studies provide information with the help of relatively short lists of quantitative measures for ranking and rating the units to be compared.

U. Teichler (✉)
International Centre for Higher Education Research Kassel (INCHER-Kassel)
University of Kassel, Kassel, Germany
e-mail: teichler@incher.uni-kassel.de

J.C. Shin et al. (eds.), *University Rankings*, The Changing Academy – The Changing Academic Profession in International Comparative Perspective 3,
DOI 10.1007/978-94-007-1116-7_13, © Springer Science+Business Media B.V. 2011

Even the exercise of defining rankings in a most simple way shows how highly normatively loaded the activities are. Are vertical differences so important compared to horizontal differences (e.g., the 'profiles' of individual institutions or units) that attention is appropriately concentrated on the vertical dimension of the higher education 'map'? Are rankings in fact instruments that disregard or even undermine the importance of horizontal diversity? Are institutional aggregates such as those of higher education institutions as a whole or their sub-units really the key carriers of quality? Are data on the quality of institutions more or less artificial aggregates of the very heterogeneous quality of academic work of individual scholars or groups of scholars within each institution which is only marginally influenced by its local environment?

Experts agree that ranking institutions is not a recent phenomenon; there is a long history of rankings. Higher education in the United States in the twentieth century has been viewed as highly successful and has served as a role model for higher education in many other countries. As well, quantitative educational measurement has been more popular in the USA than in most other countries for a long time, so it does not come as a surprise to find that many consider the USA the breeding ground of ranking studies. In fact, the first national ranking studies in the USA can be traced back to the 1920s and the first worldwide popular ranking study of universities was published in the USA in the 1980s by US News & World Report. However, there are other countries with a long tradition of national rankings. For example, rankings have played a more important role in Japan than in the USA, where competition for entry into specific Japanese universities is fiercer and where the diversity of higher education is more strongly viewed as vertically shaped with only a limited role of horizontal diversity. Consequently, there have been many ranking studies carried out in Japan over many decades. Since 1995, Asahi Shinbunsha, the publisher of the most highly regarded national newspaper, has been regularly publishing a synopsis of all available university rankings in Japan. In the 2005 edition of 'Daigaku ranking', the results of 717 ranking studies were presented.

In efforts to identify the driving forces for the emergence and spread of rankings, it is often pointed out that pervasive secular trends in higher education have been the major 'push' factors for the greater attention paid to university rankings since the 1990s. Three major trends are most often referenced:

- Massification of higher education
- Increased competition
- Internationalisation of higher education

Certainly, mass higher education, competition and international interaction must be considered in this framework. Mass higher education is viewed widely as a push factor for the stratification of higher education. When, for example, 25% of an age group graduate from higher education, the distinction between the top quintile and the second highest quintile of higher education might be functionally equivalent to the distinction between higher education and non-higher education in the past, when only 5% of students graduated from higher education. Also, the fiercer the

competition, the more attention given to the issue of whether a university is number 75 or number 80 in a ranking list. Finally, worldwide rankings are only of interest if higher education systems are not nationally segmented.

There are good reasons, however, to challenge the emphasis on secular trends when the popularity of rankings ought to be explained. In some countries, higher education had already been highly stratified before it had reached the stage of expansion commonly named 'mass higher education'. Fierce competition between higher education institutions, between students and possibly between other stakeholders had also existed in some countries before concepts of managerialism and entrepreneurialism as basic features of higher education governance at individual higher education institutions spread globally. Even with regard to 'internationalisation', there is a need to be more precise in identifying the dimensions according to which 'world-class universities' had been highly international in the past and the dimensions according to which we note an increasing internationalisation in the recent two decades.

There is another arena of discussion about the potential driving forces for the increasing attention paid to university rankings. While the arena named above is characterised by an historical analysis of the functional change of higher education, the other is shaped by psychological observations and political reflections about the paradox of rankings. It raises the question of how and why are such vague and simple measures employed to rate or rank the quality of academic work which is possibly the most complex and sophisticated feature to be assessed? Is there a hidden or overt 'virtue' in the primitiveness of information systems in higher education? How widely are the explicit or implicit ideologies of the producers of university rankings shared by the actors in higher education or its environment? Are academics so attracted by the 'excellence versus mediocrity' perspective that they consider horizontal diversity as marginally relevant at best? Do politicians believe that the quality of academic work and its relevance for society will improve if greater pressure is put on academics to follow the main stream? Does society believe in an elite knowledge society with a high concentration of academic expertise in a few locations, or does it believe in a knowledge society characterised by a spread of knowledge?

13.2 Towards a Typology of Meta-Evaluative Views of University Rankings

The discourse on the strengths and weaknesses of university rankings can be characterised as a complex interaction of methodological and functional arguments. Nobody can claim therefore that a single classification of the various types of meta-evaluative arguments clearly surpasses the quality of any other classification. In choosing a classification of arguments here, the guiding principle is to identify possible implications for the future.

The first type of argument is that the 'success story' of rankings is fundamentally based on their primitiveness. It is 'sexy' to get a quasi-objective confirmation of rumours such as conventional wisdom, surprises, new gossips, etc. Feelings of superiority and inferiority, heroism, virtue versus sin or shame, condemnation and absolution, 'excellence' and triviality are aroused. The less one knows and the less one cares about the quality of university rankings, the more one can enjoy the ranking 'games'. Of course, one has to trust blindly that there is a certain minimum authority behind the rankings. Journalistic evidence rather than academic evidence seems to suffice.

The second type of argument is based around a pragmatic discourse on the normal state of the 'quality' of indicators. Macro-societal indicators similarly defined and employed worldwide can be viewed as powerful instruments of 'transparency' and 'comparability'. We easily rely on indicators such as 'growth domestic product', 'unemployment' and 'rates of educational attainment' of the adult population. We know that there is a discrepancy between measurement with the help of a 'proxy' and the real character of the phenomenon mirrored by the indicator. We tend to accept the 'proxies' pragmatically because otherwise we would be confined to ignorance or guess work. But even if one takes such a pragmatic point of view in accepting complexity-reducing simple indicators as best possible proxies and as the 'least bad way' of measuring reality, one has to 'admit' that the rankings of 'world class universities' have not achieved the status of such general pragmatic acceptance. The expert discourse on university rankings suggests that even the simplest indicator of quality in higher education cannot be based on a single measure, but rather has to be an aggregate of various measures. Moreover, the expert discourse shows that no minimum consensus has emerged as regards a small list of key indicators to be included in an aggregate measure. However, we note a considerable readiness in higher education policy and practice as well as among higher education experts to accept the pragmatism of indicators. Citations of articles published in peer-reviewed journals are often referred to in the public discourse as a good indicator for quality of research. This often is done without any conceptual and methodological caveats that this can be interpreted as an indication that there is a readiness to accept relatively simple indicators, even in the intellectual ambitious and complexity-conscious academic environment.

The third type of argument concentrates on the assumed positive impact of rankings on higher education. 'Transparency', 'healthy competition', 'creative concentration of talents' and similar arguments are put forward.

The fourth type of argument concentrates purely on data improvement. For example: How can we reduce the problem of missing data?; How can we increase the response rate in reputation surveys?; How can we reduce institutional misallocations of authors in citation?; How can we ensure a more or less identical definition of foreign academic staff and students?; Are the definitions of staff and students similar in the statistics of the various countries and institutions?; etc. Even the argument that ranking lists should be replaced by a vertical classification of grades of quality can be viewed as a purely methodological argument. If the vertical differences between individual ranks are so small that they justify a ranking order,

a classification of ‘outstanding’, ‘very good’, ‘good’, etc. would be understood as purely methodological improvement.

Most advocates of ranking studies do not limit their critique of the current state of rankings to purely methodological weaknesses. Rather, the fifth type of argument focuses on the ‘validity’ of rankings. The term ‘validity’ is employed in the ranking discourse if one accepts the prevailing philosophy of university rankings while calling for new or improved indicators closer to the reality to be indicated. Experts who call for a better ‘validity’ of rankings often believe in or at least accept the presumed virtue of a vertical sorting of institutional aggregates in higher education with a small list of indicators as creating desirable ‘transparency’ and contributing to ‘healthy competition’. The following question addresses the need for better ‘validity’ without challenging the philosophy of rankings. How could indicators be operationalised so that they are not biased against certain disciplines, that they do not discriminate against small institutions, that they take care to strike a balance of the core functions of higher education (teaching, research and possibly service) and that they do not disregard different national conditions of higher education (for example, defining research quality not only by measuring the quality of the texts published in the English language)?

The sixth type of argument focuses on deficiencies of the prevailing ranking studies which are unlikely to be redressed in the framework of the prevailing ranking philosophies and ranking practices. The following are examples of critique that go beyond the intentions and the potentials of the prevailing ranking milieu. Rankings provide information on assumed quality differences, whereby their causes and the possible improvement remain a ‘black box’. Rankings do not take into account the ‘value-add’ achieved by the higher education institutions. Rankings claim that input, processes and output are closely linked, or that achievements in teaching and research are closely linked without taking into consideration the actual extent of linkage or dissociation. Rankings neglect horizontal diversity and are useful if an institution strives for ‘fitness of purpose’ which does not represent the main stream. Rankings claim to serve the ‘transparency’ for varied purposes, although different kind of information is needed for varied purposes: government might need information for ‘accountability’, the university management for priorities of research promotion or for strategic choices, or for the improvement of the organisational effectiveness. Students, as well, need other types of information and again other types of information are needed by possible partners of research and technology transfer. This kind of critique does not call into question the potential value of empirical data on higher education as feedback. Often, it is based on even higher expectations as regards the utility of a good data base. Rather, they consider the selection and presentation of data in the customary ranking studies as a distortion or under-utilisation of the potential of information.

The seventh type of argument might be characterised as a fundamental critique of rankings. In addition to the critique of possible biases and distortions vis-à-vis the reality of the higher education system, critique is most frequently voiced in the context of the possible or actual adverse impact ranking systems. Such impact could include undermining the extent of horizontal diversity, a discouragement of

unconventional approaches in research and teaching and an 'over-competition' which destroys potential and discourages the losers. Other consequences could be an 'over-concentration' of high quality resources in certain places that may lead to only small gains through concentration and to serious losses everywhere else. Moreover, rankings can undermine meritocratic reward by stifling the advantages of the historically privileged institutions and of the winners due to symbolic advantages and successful short-term tactics. Finally, rankings might mobilise the above-average institutions while having a zero-effect or even discourage below-average institutions which need appropriate feedback in order to improve. Whatever the distortion of the data, they are likely to elicit even higher distortions as a vicious circle of mal-information and adaptive behaviour. As a rule, experts voicing such a fundamental critique of rankings leave it open as to whether they believe in possible improvements of systematic empirical information as feedback for higher education, or whether they consider such efforts as futile because rankings are bound to produce distorted information.

13.3 The Possible Futures of Rankings and Beyond

We have experienced many rapid changes in higher education over the last few decades, and many changes were not predicted beforehand. As a result we do not feel confident in predicting the future as far as university rankings are concerned. But we can suggest some likely scenarios.

First, we may experience a situation best described as the inertia scenario. If an element of higher education has been present for quite a while or has emerged in recent years, it is 'here to stay', as many advocates of university rankings point out. Interest in vertical lists of universities and a belief in their quality and virtues will be too stable to challenge them, and there is no evidence that this feature will be not protected by the widespread system inertia in general.

The second type of possible scenario could be the trend scenario. In several countries, there is a long tradition of rankings. In recent years, interest in rankings has spread to other countries. As ranking construction increases and greater attention is paid to rankings worldwide, this trend will be reinforced by other trends such as massification, increasing competition and internationalisation. As a result, a further spread of rankings can be expected.

Third, we suggest a 'politics works' scenario. Rankings may change higher education that the way the ideologists of the ranking movement hope and primary critics of higher education rankings fear. For example, horizontal diversity may become more or less irrelevant for higher education, and the competition for the highest possible rank according to relatively homogeneous criteria might become even more pervasive.

Fourth, we can imagine an emerging problem and emerging problem awareness scenario. For example, if main stream rankings are biased towards the research functions, there is potential for serious problems in the quality of teaching and learning.

This in turn may lead to major steps being taken to redress this deficiency, including a reform of an information system which tends to elicit undesirable adaptations.

Fifth, it is also worth considering alternative scenarios where a more desirable higher education system emerges. The recent spread of 'diversity management' is one example of newly emerging paradigms which could challenge the basic concepts underlying rankings. In this scenario, inter-institutional (vertical) diversity concurrent with a relatively high intra-institutional homogeneity would become the most desirable and productive and the idea might spread that intra-institutional diversity will be a matter of fact and will be the most productive future of higher education.

Sixth, there is no reason to exclude the potential for a 'turn towards high quality rankings' scenario. The methodological optimists may turn out to be right in saying that those who are involved in the production of rankings and those who are involved in the funding of rankings or of other information systems on which rankings are based, are willing to strive for a higher complexity of rankings, as well as for a broader concept of 'validity' than those which are now in place.

Seventh, at least for the sake of logical completeness, we suggest the 'increasing complexity of balanced information' scenario. Open, thorough and unbiased feedbacks, as well as evidenced-based strategic action may become so highly appreciated in a knowledge society that biased systems of information gathering can no longer overshadow other relevant information systems. What role would remain for rankings if a scenario of impressive transparency and rationale actors became a reality?

Index

J.C. Shin et al. (eds.), *University Rankings*, The Changing Academy – The Changing Academic Profession in International Comparative Perspective 3,
DOI 10.1007/978-94-007-1116-7, © Springer Science+Business Media B.V. 2011