This is the accepted version of a paper published in *Quality in Higher Education*. This paper has been peer-reviewed but does not include the final publisher proof-corrections or journal pagination.

Citation for the original published paper (version of record):

Hägg, I., Wedlin, L. (2013)
Standards for quality?: A critical appraisal of the Berlin Principles for international rankings of universities.
*Quality in Higher Education*, 19(3): 326-342
http://dx.doi.org/10.1080/13538322.2013.852708

Access to the published version may require subscription.

N.B. When citing this work, cite the original published paper.

Permanent link to this version:
http://urn.kb.se/resolve?urn=urn:nbn:se:uu:diva-211355
Standards for quality?

A critical appraisal of the Berlin Principles for international rankings of universities

By Ingemund Hägg and Linda Wedlin
Department of Business Studies
Uppsala University
Box 513
751 20 Uppsala
E-mail: Ingemund.hagg@fek.uu.se, Linda.wedlin@fek.uu.se

Abstract
This paper discusses the principles developed to assure the quality of international ranking practices for higher education, the so called Berlin Principles, and the role given to them in the higher education community. While the principles are generally regarded as proper quality assurance principles, we argue that they are problematic both in their content and form. By examining the process leading up to the principles as well as their use, we underline some legitimacy principles with these principles and the potential impact this may have on the further development of the higher education field. The study is based on official documents and published material.
Standards for quality?

A critical appraisal of the Berlin Principles for international rankings of universities

Abstract

This paper discusses the principles developed to assure the quality of international ranking practices for higher education, the so-called Berlin Principles, and the role given to them in the higher education community. While the principles are generally regarded as proper quality assurance principles, we argue that they are problematic both in their content and form. By examining the process leading up to the principles as well as their use, we underline some legitimacy principles with these principles and the potential impact this may have on the further development of the higher education field. The study is based on official documents and published material.

Establishing principles for rankings

International rankings have become a core feature of the competitive landscape of universities and higher education institutions. These rankings proclaim to provide students and other interested stakeholders with information and guides to value for money of educational offerings, and to provide universities and higher education institutions with benchmarking standards and assessments of the relative standing and competitive position of its programs and the institution. Such rankings have proliferated over the past decade, and increased in number as well as in scope. That these rankings have a significant impact on the higher education landscape is by now well-documented (Hazelkorn, 2007; 2008; Kehm & Stensaker, 2009; King 2009; Locke et al., 2008; Wedlin, 2006).

The issue for this paper is not on the impact of rankings, but on the legitimacy and development of the rankings as measurement practices. How have rankings established their positions as influential measures of performance of higher education organizations across the world? By investigating one of the recent efforts to establish criteria on which to judge and evaluate ranking practices, we seek to increase understanding of how ranking practices become legitimate elements of the global higher education landscape.

With the proliferation of international rankings of universities over the past decade have come concerns for how these rankings are constructed, what they measure and, more recently, which of them are trustworthy and reliable measures of performance and relative standing of world universities (Harvey, 2008; Stolz, Hendel & Horn, 2010). In part as a response to such concerns rankers and others involved in them have joined forces to develop a set of guidelines for developing and assessing the quality of different rankings and their measures, called the Berlin Principles (BP). These principles, as 16 general guidelines concerning the data collection, methodology, dissemination of rankings and their assessments, have recently also been used as the foundation for an audit practice of university rankings, the IREG Ranking Audit, launched in 2011. This audit, performed by the IREG International Observatory on Academic Ranking and Excellence, serves as a certification of rankings, by providing the quality seal of ‘IREG approved’ to rankings that passes the assessment.

While the influence of this development still is to be seen, we believe it is time for a critical reflection and discussion of the BP and its role in shaping the influence of rankings and ranking practices – and thus subsequently the definition and assessment of
quality and worth – in the field of higher education. The attention that so far has been paid to these principles has to a large extent celebrated the principles as a great leap forward in the development and evaluation of rankings and in the forming of appropriate measures of academic and particularly university performance and relative standing (Chen & Liu, 2008; McCormick, 2008). But are these principles really setting relevant standards for assessing or guaranteeing the quality of academic rankings? If so, how do they do that, and what role do they play in shaping the agenda for higher education quality and assessment work?

In this article we provide a critical analysis and discussion of three aspects of the Berlin Principles and its practices: How and by whom the BP are constructed, what they contain, and how they are being used. We will relate our discussion to an accountability framework with the focus upon dilemmas of accountability presented by Peter Miller (1996). He introduces three dilemmas. “First, the dilemma of principle: should we place our trust in experts?” In our case – trust in ranking and trust in the BP. “Secondly, the dilemma of meaning” where Miller focusses on the interpretation made by users and asks the questions of what it means to be a “customer.” In our case what it means to be a user of rankings and of ranking principles. “Thirdly, the dilemma of mechanism.” Miller poses this dilemma with the question “can we place our trust in numbers.” We will discuss and use this framework in our analyses of the role of the BP. Based on this we will discuss the potential impact the BP have on the field of higher education.

**Rankings for accountability**

International rankings for higher education, mainly for universities, have developed and proliferated rapidly since the early 2000s. The two most known rankings are the Academic Ranking of World Universities (ARWU) produced by the Shanghai Ranking Consultancy in China, and the World University Ranking produced by the British newspaper *Times Higher Education* (THE). These have been produced annually since 2003 and 2004, respectively. Since then initiatives to rank and make competitive league tables and comparisons of universities worldwide have flourished.

While not the first rankings of universities to appear, the ARWU and THE rankings have had a significant impact in the field of international higher education. The rankings have gained an increasing interest within and among universities around the world, and universities in varied contexts became interested in comparisons and assessments of themselves and others. At conferences, meetings and in networks, the issue of rankings has been a prominent feature of university debates in recent years. Focused conferences dealing with rankings have been organised in for instance Shanghai in 2007 and Leiden in 2009, but also elsewhere. The ranking debate has also spurred other actors, such as governments and international organisations, to pay attention to rankings (Hazelkorn, 2007; King, 2009). A particular example is the EU effort to develop a new, comprehensive multi-dimensional ranking of world universities (U-Multirank) (SeeVan Vught & Westerheijden, 2010).

We can interpret the proliferation of and increasing interest in rankings in many ways. First, it is an element in a change in the governance practices of the higher education field in recent years. Such governance changes include a decrease in state regulations, and an increasing concern for quality assurance and assessment practices, both among states and other actors in the higher education field (Hedmo et al., 2006; King, 2009). This is seen as a growth in accreditation practices and schemes for higher education (Hedmo, 2004; Schwarz & Westerheijden, 2004). Second, in recent years we have witnessed an increasing political and public interest in universities as “knowledge-producing” organizations, aimed to provide benefits, mostly economical, for societies, nations and regions (Frank & Meyer, 2007). This has created demand for universities and higher education organizations to be accountable, to
the state and to the public, for the use of public resources and to meet expectations of usefulness and value in society. Thus, we have witnessed a significant growth in and development of accountability mechanisms and organisations for higher education in many countries (Stensaker & Harvey, 2011). Third, higher education institutions are increasingly seeking visibility on what is perceived as an international market. This is tied, in part, to developing international comparisons and accreditation procedures, but also to a general interest in reputation and image. Ranking have, in this context, become important mechanisms to create the appearance of a global, or a unified, field of higher education (Dill & Soo, 2005; King, 2009).

Together, these three developments suggest that we have a partly new set of demands, expectations and situations for universities and other higher education organizations to handle. These include, but are not restricted to, demands for accountability and transparency in organizational practices and procedures, and expectations of similarity and comparability across countries and regions. It is in the spirit of these demands that rankings and other external evaluation and monitoring mechanisms develop and gain prominence (cf. Power, 1997; Miller, 1996).

An initiative to increase legitimacy

The proliferation and increasingly prominent role of rankings in the international university debate has led to a number of efforts to organise and to support the development and improvement of rankings across the world. Among the most prominent efforts of this kind is the forming of the International Rankings Expert Group (IREG) in 2004. IREG was formed on the initiative and with the support of the UNESCO European Centre for Higher Education (UNESCO-CEPES) in Bucharest and the Institute for Higher Education Policy (IHEP) in Washington D.C. It met for their first ranking conference in Washington in 2004. Since then, this group has arranged regular yearly or biennial ranking conferences in different places around the world to discuss rankings and ranking practices. In 2006 they launched the Berlin Principles on rankings of higher education institutions (BP).

Before we turn to the details of the BP to discuss their importance for this field it is important to understand the driving forces and the motives of these initiatives and how these became elements of the global ranking field. We can see the organising of IREG and the launch of the BP as elements of an effort to establish and increase legitimacy for ranking practices in this field. This will be evident as we briefly review the process leading up to the organising of IREG and the launch of the BP.

A self-organised initiative

The organising of IREG was initiated by the UNESCO-CEPES and the IHEP in the early 2000s. The initiative largely grew out of an interest within these organisations for performance evaluations and measurement systems for higher education and research more generally. As a prelude to the forming of IREG was a series of roundtables and conferences organised by UNESCO-CEPES between 2001 and 2003 for a project entitled ‘Strategic indicators for Monitoring Higher Education in the Twenty-First Century’ (UNESCO-CEPES 2002). This project originated in the work of this organisation to initiate and support a reorientation of higher education systems and institutions around the world, particularly perhaps throughout its member states in Central, Eastern and South East-Europe (UNESCO-CEPES 2011). In a statement following a 2002 roundtable, the group identifies rankings and league tables as particularly important to bring about change in the higher education sector. They write:
In order to bring about the kind of renewal and reorientation required, policy-makers and administrators need a quantitative, solid system of fact-reporting, applicable even to daunting cases, so as to be able to quantify the intangibles of a set of complex teaching, learning and research phenomena, and the administration, functioning and financing of higher education. (Unesco-Cepes, 2002, p. 359)

Based on this work, initiative was taken for the 2004 Washington meeting, at which IREG was constituted. This meeting was attended by 25 people interested in and in different ways involved in practices and organisations related to quality assurance and rankings in higher education. Among these, we find representatives from the rankers THES, ARWU and the US News and World Report, as well as the German centre for higher education, the CHE (Centrum für Hochschulentwicklung), developing a ranking system published by the magazine Stern. Organisations such as the German Research Foundation (DFG), the Austrian Agency for Quality Assurance, and the Educational Policy Institute in Canada also had representatives at this meeting. RAND Corporation also participated. Remaining participants were researchers and administrators from universities in the US, Russia, Japan, Poland, China, and Spain.

The rationale for the 2004 meeting was that rankings were noted to play an increasing role as information tools for consumers, and that IREG would play to the "need for informed international debate with regard to the phenomena of institutional and program ranking in higher education" (IREG, 2004, p.1). The specific purpose of the group was to provide research on rankings, contribute to framework to assess and compare existing rankings and their methodologies, and ultimately to "improve existing ranking/league tables and inform future ranking methodologies" (IREG, 2004, p. 2). An explicit aim was however also to legitimise ranking practices – which were rather highly contested at this time – within the broader higher education field. Jan Sadlak, Director of UNESCO-CEPES and one of the founding members of IREG, thus stated that this group had "the potential to extend credibility to the rankings/league tables and to ensure that they more effectively reflect higher education quality" (IREG, 2004, p. 2).

The IREG meeting in Washington was followed by a second meeting arranged in Berlin in 2006, co-hosted with the German quality agency CHE. This meeting held the title Methodology and Quality Standards of Rankings, and was specifically aimed to discuss and develop guidelines for ranking practices across the world. James Merisotis, president of IHEP, here presented the Berlin Criteria and Standards for International University Ranking, partly elaborating on the findings and reflections after the previous IREG meeting (see HEiE 2005). These standards, after some revision, were also discussed and subsequently adopted as the Berlin Principles on Ranking Higher Education Institutions by the group at this meeting (IREG, 2006).

Rank others, certify yourself
Following the establishment of the BP, IREG continued its work to promote and discuss rankings at various meetings and conferences for higher education institutions and organisations across the world. At their next conference, the third IREG meeting held in Shanghai in 2007 (co-hosted by the Shanghai Jiao Tong University), the group began to discuss ways to formalise and structure its continued work on university rankings in new ways. An idea to launch an Observatory for the rankings was introduced. This Observatory would be a more permanent organisation responsible for the continued work to promote and improve ranking practices throughout the world.

On April 18, 2008, the IREG Observatory on Academic Ranking and Excellence was established. With headquarters in Warsaw, Poland and secretarial assistance provided by
the Polish organisation Perspektywy Education Foundation (among other things publisher of the Polish ranking for the magazine *Perspektywy*), the organisation continued the work of IREG with many of its most noted members now on the executive committee. Among these we find president Jan Sadlak (formerly director of UNESCO-CEPES), representatives from ranking organisations (ARWU, CHE, *Perspektywy*, Slovak Academic Ranking and Rating Agency (ARRA), the Independent Kazakhstan Quality Assurance Agency in Education (IQAA.KZ), and *US News and World Report*), as well as organisations such as IHEP and the Educational Policy Institute in Toronto, Canada.

The Observatory was seen to be a strategic re-orientation of the work of IREG, and a ‘formula for establishing academic ranking as a credible instrument of qualitative performance for higher education and the institutional and program levels’ (IREG, 2009, p. 9). Apart from continuing to initiate and present research and debate on rankings and ranking practices, the Observatory takes as an explicit aim to work at ‘improving the standards, theory and practice [of rankings] in line with the Berlin Principles of Ranking of Higher Education Institutions’ (IREG, 2009, p. 9). In fact, to implement and to diffuse the BP was the stated reason to formalise IREG into the Observatory (IREG, 2010).

The objective of the Observatory to diffuse and implement the BP has been further elaborated by the work to establish a ranking audit, or a certification procedure for rankings and ranking organisations across the world. Thus, at the fifth IREG conference in Berlin in 2010 the new IREG Ranking Audit was presented and launched. Significantly based on the BP, the audit is expected to:

- enhance the transparency about rankings; give users of rankings a tool to identify trustworthy rankings; and improve the quality of rankings (Sadlak, 2010, p.2).

Following consultation from participants at the Berlin conference and on the website during the fall of 2010, the final ranking audit rules and procedures were adopted by the IREG Observatory executive committee in May 2011. The first audits will be based on applications made before the end of January 2012. In May 2013, two rankings were the first to be granted the “IREG approved” label; the Polish Perspektywy University Ranking and the international QS World University Ranking.

The audits are conducted by an Audit Team consisting of three to five members appointed by the Executive committee of the Observatory, which will also set the audit fee for each audit. At least one member of the Audit Team has to be a member of the executive committee of the Observatory. The chair of the Audit Team may not be formally associated with a ranking organisation. The Executive committee makes the final decision on the audit, based on recommendations from the Audit Team and statements from the ranking organisation being audited. Positive audits will give the ranking organisation the label 'IREG approved’ and the names will be published on the IREG website (IREG, 2011). IREG has decided that only positive audit decisions will be made public, and that individual audit scores will not be published at any time.

**Principles for legitimacy**

Being part of the efforts of rankers to legitimate the practices and procedures of rankings in the field of higher education, the BP serve as a framework to evaluate and assess ranking systems in various contexts and settings. So how do they do this? While the BP provide general guidelines rather than specific recommendations, it is possible to identify five basic issues being addressed in these principles: clarity, transparency, completeness, relevance and legitimacy. Below we will briefly show how the 16 different principles relate to these issues, and thereafter discuss two significant problems inherent in this set of guidelines.
Several of the BP relate to and discuss issues of clarity in the rankings, and the importance of being clear and precise in their reports (BP number 2, 4, 9 and 15, see table 1). These state that rankings should be clear in purpose and target group, to which the rankings are aiming, and in the range of information provided. Weightings of different criteria and aspects should also be clear and stable over time, and they should provide a clear understanding of all the factors used in the final ranking.

Related to clarity is also the issue of transparency (presented in BP number 3, 5, 6, and 11). Transparency is about the need to present audiences with the opportunity to evaluate and understand the contexts, and limitations, of the rankings. Clarity cannot automatically be assumed to provide transparency in this sense. Transparency is illustrated in the principles that state that rankings should recognise the diversity of organisations being evaluated and be able to specify the varying contexts in which these organisations operate. Furthermore, the methodology should be clear and unambiguous, and the calculation of indicators as well as the origin of data should be transparent. Audited and verifiable data should be used where possible.

Completeness relates to the demand that rankings should be as complete and comprehensive as possible (BP number 4, and 15). This is expressed by the demand that rankings combine different perspectives by combining different data sources (databases, student surveys, faculty surveys). Rankings should also be complete in the sense of providing users a choice in how the rankings are displayed, allowing them to make their own decisions as to the weight and relative importance of different indicators. While being rather ambiguous, BP number 8 also state rankings should include both input measures and output measures in order to provide a more balanced view of higher education organisations.

Three of the principles also relate to the issue of relevance (BP number 4, 7 and 8). Issues of relevance are touched upon rather briefly in most of these cases, such that it is stated that the relevance of the rankings depends on the audiences receiving the information and the sources of that information (BP number 4). While input measures are considered relevant in some cases, one principle states that measures of outcomes are preferable (BP number 8). Only one principle explicitly argue that indicators should be chosen based on their relevance and validity, and thus on their ability to ‘represent quality and academic and institutional strengths’ (BP number 7).

The final issue deals with the legitimacy of the rankings. Several of the principles are more or less specifically aimed to secure or improve the legitimacy and credibility of specific rankings as well as of ranking practices in general, expressing that it is necessary to ‘assure the credibility of each ranking’ (BP number 10), and, more generally, to ’enhance credibility’ (BP number 14). Several principles touch upon this issue (BP number 9, 10, 12, 13, 14, and 16). This is expressed in demands that weights assigned to different indicators and criteria should be made prominent and changes to them should be limited. Ethical standards and good practices should be applied in collecting and using data, particularly stressing the need to be objective and impartial in such processes. Furthermore, ‘proper procedures’ should be applied to data collection procedures, and measures of ’quality assurance’ should be applied to the rankings themselves. Credibility of rankings is to be furthered also by including so called organisational measures to develop the rankings, such as advisory or supervisory bodies that can oversee the rankings process. The final principle suggests that the final rankings should be compiled and presented in a way that eliminates or reduces errors in data, or informs the public where such errors have occurred.

Contradicting demands and expectations
A first note we can make about these guidelines is that there are a number of inconsistencies and seemingly contradicting demands and expectations put on rankings through these
principles, both within and between them. While we do not make an exhaustive analysis of these here, we will point out those that seem to make interpretations of the guidelines problematic.

First, the arguments for clarity seem to sit uneasily with some of the criteria for transparency, or, at least, it is not certain that the latter comes with the former. Clarity, stating that the purpose and target groups of rankings should be clear, seems reasonable but problematic if it is to be combined with the need to recognise diversity of higher education organisations. To be transparent and clear about the diversity of missions and goals among higher education organisations seems to require a more elaborate discussion about the role and purpose of rankings than what can be drawn from one or a few defined target groups. Furthermore, the need to make transparent and understandable the ‘linguistic, cultural, economic and historical contexts’ of different educational systems presupposes a system of well-defined and identifiable classes or groups of institutions that can be compared, which require additional information than what is possible to provide even within a ‘clear and transparent’ ranking system.

Clarity, and also transparency, also requires a more nuanced view of the possibility, and aim, to create more complete ranking systems. While clarity should be provided about all information sources and its uses, combining many different sources that provide complementary, or contradictory, views of the ranked institutions may not contribute to make the rankings clearer or easier to interpret. Furthermore, a combination of different sources may not automatically lead to a more complete ranking, as this depends on the choice of sources and how this information is presented and used to supplement or contradict other findings. This is suggested in the principles about relevance, which further seem to confuse the message about both clarity and completeness.

The aim at completeness, both in terms of combining perspectives and data sources and by including both input and output measures, is also contradictory to the idea of relevance. Here it is stated that output measures are more relevant than input measures, and that criteria should be selected based on relevance for quality and academic strength. These are, however, not connected, and a balance between input and output measures is implicitly assumed to be preferable to an unbalanced weighting, regardless of its relevance for quality. This may also, in principle as well as in practice, contradict the ideal of transparency as using audited and verifiable data, which may not always be either accessible or particularly relevant data.

On a more general note a contradiction lies in the very message of the principles itself: to continuously improve the rankings and contribute to a better ranking practice overall, thus to make the rankings ‘learning systems’ that develop in collaboration with the evolving expertise in rankings (BP number 13). At the same time, a fundamental principle in the construction of indicators and weightings is that these should be kept stable and that changes to them should be limited in order to assure comparability over time (BP number 9).

**Problematic assumptions**

While some of the principles may place contradictory demands on the rankings, many of the guidelines are rather straightforward and place clear demands on the rankings to produce valid and reliable measures that are relevant for assessing quality. This is reasonable, but holds another problem inherent in these kinds of measuring systems; that is, it tends to build on assumptions about what quality is and how quality can be assessed. The principles, in this sense, are no exception, and they rest on several assumptions about both the availability of data and the acceptance of measures of quality within the area of higher education.

Given the attention in the principles about the need to situate rankings in a context, and provide information and data that is relevant and understandable to specific
audiences, it is striking to note the lack of reflection of the role and purpose of rankings in relation to the wider quality debate in higher education and research. Indicators discussed in connection with rankings lack, of what we can see, theoretical bases and are rather of ad hoc character. The choice of weights of indicators is not discussed in connection with any kind of theory and thus seems arbitrary, chosen rather subjectively by the ranker (for example BP number 9). This implicitly assumes that rankings are robust for changes in weightings. It also assumes that information usefulness for target groups can fruitfully be assessed by the producers of a ranking. The principles, for example, do not deal with issues of interactivity between users and rankers, apart from offering a choice as to how data is presented (BP number 15).

A fundamental assumption is that rankings overall provide a meaningful set of comparisons and that it is, in principle as well as practice, possible to establish a fair, acceptable and ‘true’ methodology for rankings. While the principles suggest that ranking should respect diversity of higher learning institutions in several important respects, it is still unclear how this should in fact be represented and incorporated into a single ranking list. It is rather assumed that it is possible to keep the number of classes of institutions low, which means that some diversity is tolerated and assumed not to distort a specific ranking.

**Ranking the rankings**

In the above section, we have discussed some problems inherent in the recommendations and guidelines provided by the BP. Many of these are not, however, specific to this particular set of principles and guidelines, but characterise many of such systems to measure, evaluate or set standards for activities and organisations (Miller, 1996). This may not be a problem, however, depending on how such guidelines and standards are being used. We will in the following briefly discuss some of the ways these principles have been used to assess or evaluate rankings and what this may mean for the practice of ranking.

The debate following the launch of the BP has been mainly positive about the usefulness and positive contribution of the BP to the general debate about rankings. Proponents of rankings, rankers themselves, as well as members of the IREG group have made significant effort to launch the BP as the ‘golden standard’ for rankings, not least through academic articles and contributions to international higher education journals (most prominently the journal Higher Education in Europe, until 2009 published on behalf of UNESCO-CEPES). Cheng and Liu (2008), for instance, see the BP as ‘a milestone in academic rankings’. They furthermore see the IREG expert group as the best forum for developing the principles further (of which they themselves are a part). Stolz et al. (2010) are no less positive to the expert group which ‘is in a unique position to discuss and present such examples [i.e. practical guidance]’. Neither Chen and Liu, nor Stolz et al. show signs of criticism of the work of the expert group. The group and the principles have also garnered wide support from the academic community more widely, and in a report from the European University Association in 2011, entitled Global university rankings and their impact, the author concludes that ‘[t]here is no doubt that the Berlin principles are a good guide and ranking institutions often claim that they comply with them’ (Rauhvargers, 2011: 21).

Some of the writings on the BP have used them to evaluate or compare different rankings, thus a kind of ranking of rankings. Cheng and Liu (2008) transform the principles into criteria which they then use for evaluating how eight selected rankings conform to the BP. They develop 14 criteria which refer to 13 of the BP. For three of the principles they have not been able to find measurable criteria (principles 7, 8 and 11). One of the conclusions drawn by Cheng & Liu (2008) is that there are great problems in transforming the BP into measurable criteria. Most significantly, this applies to BP number 5 and 3 that deal with contextualisation of international rankings and the need to specify and take account of diverse
missions of higher education institutions. Regarding the first, they conclude that ‘if an international ranking used different indicators to compare the HEIs [higher education institutions] of different countries, international rankings would be meaningless’. As regarding the second, they conclude that there are attempts to deal with diverse institutions but that there is a lack of clear classification systems that can be used, and that there simply are no good solutions yet out there that accurately take account of such diversity.

In a similar manner Stolz et al. (2010) rank 25 higher education institutions with the help of 14 quantitative measures derived from the BP. Some of the measures are divided into two sub-scales. For their study they measure ten of the 16 principles, excluding principles 1, 5, 10, 13, 14 and 16. The reasons for excluding these are primarily that they lie outside the scope of assessing the ‘quality of ranking practices’ (Stolz et al., 2010, pages 511 and 514), and that there is a significant lack of reliable information on issues of ethical standards, quality assurance procedures, or error reducing practices of the rankings. They furthermore note some measuring problems with BP number 7, 9 and 12. They also note, however, that eliminating many of these measures ‘might lessen the meaningfulness of the ranking of HERSs [higher education ranking systems] as undertaken in this study’ (Stolz et al., 2010, p. 517).

In applying the measures elaborated on, Stolz et al. end up with a ranking of rankings, scoring the ranking systems on the various measures derived from the BP. They note in this computation, though, that no ranking in their sample of European rankings ‘achieves a good overall congruence with the best practice proposed by the BPs’ (Stolz et al., 2010, p. 521). They mainly point to criteria related to methodology as particularly weak in this respect. One of their conclusions drawn from this finding is that there is a significant lack of guidance on how rankings actually can conform to the BP, and that the BP therefore are rather difficult to put into concrete practice for the rankers. Particular problems are found with the criteria for relevance and validity, where they note that the ‘expert group has so far provided neither a theoretical rationale nor a practical guideline for achieving validity’ (Stolz et al., 2010, p. 515).

Using our categorisation of the different ranking criteria into the five main groups – clarity, transparency, completeness, relevance, and legitimacy – we can conclude from the above discussion that some of these seem to be more problematic than others. While the principles for clarity all seem to be reasonably easy to translate into criteria with which to assess rankings, there is significant more problems with the principles relating to transparency, relevance and, perhaps most pronounced, with criteria relating to legitimacy of the rankings when turning the principles into detailed measures. Thus, we can conclude that the BP are problematic to actually use for assessing or comparing existing ranking practices.

**Setting standards for quality?**

Our review above has shown that the work to establish the BP can best be understood as a process to assure the legitimacy and primacy of the rankings as measures of performance of higher education organisations. The work leading up to the creation of IREG, the establishment of the BP, and the subsequent creation of the ranking audit all point essentially to this need to legitimise the ranking practices and the rankers themselves. The self-appointed expert group IREG is constituted to a large extent by these ranking organisations and their representatives, and they are a clear driving force in this work.

The problems with the BP as discussed above also highlight some more general characteristics and problems with measurement systems and audit/accountability systems. Thus situating this critique in the wider debate about the limits of such measuring and accountability techniques – particularly accounting and auditing practices – we will relate our analysis to the framework we briefly presented in the beginning of the article, that is the three
“dilemmas of accountability” developed by Miller (1996). These are dilemmas of principle, meaning and mechanism.

The first of these dilemmas is that of principle: can we place trust in experts? The increasing demand for accountability witnessed in recent years is driven in part by distrust in the higher education field and its experts, and thus a lack of trust in the internal academic evaluation and quality assurance principles. Supporting this, the rankers are forcefully arguing for the need of rankings because there are customers, clients, and other stakeholders to whom higher education is and should be held accountable. With this, however, the localisation of trust is shifting place: trust seems to have moved to rankers and other external actors to judge and to evaluate quality, performance and value in the higher education field, at least judging from the increasing salience of these mechanisms in the university governance debate.

In light of our analysis of the BP here, we can also extend this dilemma further, and ask: which experts can we trust? Given the rather substantial debate and controversy about rankings and the aspects of the accountability they attempt to provide, questions about trust become again salient. How is accountability to be defined, to whom is accountability to be held, and who can we trust to provide relevant and trustworthy information? Our analysis of the BP suggest that as questions of legitimacy and authority of rankings and ranking organizations are being raised within the higher education field, we are now asked to put trust in yet other mechanisms and actors to assure that the rankings can be trusted: in the form of guidelines, organizations, and audit systems (the BP, IREG and the IREG audits). As foreseen by Power (1997) and others, audit practices tend to produce not less but rather continually more auditing and a spiral of ever increasing demands for trust and accountability.

Our discussion of the BP illustrates further the problems of this shifting localisation of trust. We have shown how the IREG group has named themselves as experts in the field of evaluating and judging the performance and quality of rankings. We argue that there are significant problems with the authority and legitimacy of this group, as well as with the principles and practices of evaluating ranking that they present. Some rankers have acknowledged this, refusing to participate in or recognise the authority of this group to evaluate ranking practices.

Our case also illustrate the two other dilemmas of accountability, namely that of meaning and mechanism. The dilemma of meaning suggests that there are problems with interpreting and judging what it is exactly that is being assessed in these systems, and what that implies. For users of rankings, it is unclear what the principles established in the BP represent, and how they help to assure the quality of the ranking systems which they assess. The dilemma of mechanism suggests that there are also practical problems in assuring accountability through the means of these systems, mainly through the use of numbers. Most significantly, we have shown how there are difficulties in interpreting and clearly see what the BP represent and further, even greater problems to put the standards and guidelines into practice. The BP rely to a great extent on assumptions of both measurability and quality which are at the same time contested and vague.

Concluding remarks
Our review of the Berlin Principles for rankings has focused on three particular issues; how and by whom the BP are constructed, what they contain, and how they are being used. Regarding the first issue, we can note that the principles have been developed largely within a self-organised network of actors that have an interest in rankings and ranking practices. Many of the most significant actors in this network are themselves rankers, or representatives of organisations that produce and/or market rankings in the field of higher education and research. The second issue concerns what the principles contain and what they attempt to
measure, and we here note five underlying categories for these principles: clarity, transparency, relevance, completeness and legitimacy. Many of these are both ambiguous and contradictory, and reflect some underlying assumptions about the quality and applicability of rankings overall. The third issue points to the uncritical way in which these principles have been used, mainly to attempt to rank the rankings. These attempts, however, rather point to measuring and applicability problems in using the principles to assess or improve the quality of rankings.

Based on this review we can conclude that the BP and the work of the IREG group to establish standards and principles for international rankings can best be understood as a self-organized initiative aimed to create and enhance legitimacy for ranking practices overall. As there are significant problems both with the authority of the IREG itself – being a self-appointed expert group – and with the interpretation and use of the principles themselves, it is difficult to find the BP of any real value to either assess, compare or improve ranking practices in any fundamental way. In contrast to those considering the BP to be the ultimate solution to the problem of rankings, we conclude that the Berlin principles were a worthwhile contribution to the discussion of ranking of higher educational institutions, but little more. Highlighting some of the problems and inconsistencies with these types of systems thus teaches us to ask more fundamental and ultimately relevant questions about the diversity, quality, and validity of higher education and its assessments.

Our focus in this paper has been on the role of the BP, and we have not aimed at discussing alternatives or complements to existing ranking practices. While our own view is that rankings imply reducing information to numbers and ranks and that there is a need to develop information systems that can provide “deeper” information with relevance to users, this is not the place to develop these thoughts further.
Table 1. Berlin Principles

<table>
<thead>
<tr>
<th>Category</th>
<th>Principle</th>
</tr>
</thead>
<tbody>
<tr>
<td>CLARITY</td>
<td>2. Be clear about their [rankings] purpose and their target groups.</td>
</tr>
<tr>
<td></td>
<td>4. Provide clarity about the range of information sources for rankings and the messages each source generates.</td>
</tr>
<tr>
<td></td>
<td>9. Make the weights assigned to different indicators (if used) prominent and limit changes to them.</td>
</tr>
<tr>
<td></td>
<td>15. Provide consumers with a clear understanding of all of the factors used to develop a ranking, and offer them a choice in how rankings are displayed.</td>
</tr>
<tr>
<td>TRANSPARENCY</td>
<td>3. Recognize the diversity of institutions and take the different missions and goals of institutions into account.</td>
</tr>
<tr>
<td></td>
<td>5. Specify the linguistic, cultural, economic, and historical contexts of the educational systems being ranked.</td>
</tr>
<tr>
<td></td>
<td>6. Be transparent regarding the methodology used for creating the rankings.</td>
</tr>
<tr>
<td></td>
<td>11. Use audited and verifiable data whenever possible.</td>
</tr>
<tr>
<td>COMPLETENESS</td>
<td>4. Provide clarity about the range of information sources for rankings and the messages each source generates.</td>
</tr>
<tr>
<td></td>
<td>15. Provide consumers with a clear understanding of all of the factors used to develop a ranking, and offer them a choice in how rankings are displayed.</td>
</tr>
<tr>
<td>RELEVANCE</td>
<td>7. Choose indicators according to their relevance and validity.</td>
</tr>
<tr>
<td></td>
<td>8. Measure outcomes in preference to inputs whenever possible.</td>
</tr>
<tr>
<td></td>
<td>4. Provide clarity about the range of information sources for rankings and the messages each source generates.</td>
</tr>
<tr>
<td>LEGITIMACY</td>
<td>[1. Be one of a number of diverse approaches to the assessment of higher education inputs, processes, and outputs.]</td>
</tr>
<tr>
<td></td>
<td>9. Make the weights assigned to different indicators (if used) prominent and limit changes to them.</td>
</tr>
<tr>
<td></td>
<td>10. Pay due attention to ethical standards and the good practice recommendations articulated in these Principles.</td>
</tr>
<tr>
<td></td>
<td>12. Include data that are collected with proper procedures for scientific data collection.</td>
</tr>
<tr>
<td></td>
<td>13. Apply measures of quality assurance to ranking processes themselves.</td>
</tr>
<tr>
<td></td>
<td>14. Apply organizational measures that enhance the credibility of rankings.</td>
</tr>
</tbody>
</table>
16. Be compiled in a way that eliminates or reduces errors in original data, and be organized and published in a way that errors and faults can be corrected.

References


Hedmo, T., 2004, Rule-making in the transnational space: The development of European accreditation in management education. (Uppsala University, Uppsala).


