Open Research Online

The Open University's repository of research publications
and other research outputs

Ranking Meets Distance Education: Defining Relevant Criteria and Indicators for Online Universities

Journal Item

For guidance on citations see FAQs.

Version: Version of Record

Link(s) to article on publisher's website:
http://dx.doi.org/doi:10.19173/irrodl.v20i5.4391
http://www.irrodl.org/index.php/irrodl/article/view/4391

oro.open.ac.uk

Ranking Meets Distance Education: Defining Relevant Criteria and Indicators for Online Universities

Francesca Pozzi[1], Flavio Manganello[2], Marcello Passarelli[3], Donatella Persico[4], Andrew Brasher[5], Wayne Holmes[6], Denise Whitelock[7], and Albert Sangrà[8]
[1,2,3,4]*Institute of Educational Technology - CNR, Italy,* [5,6,7]*The Open University, UK,* [8]*Universitat Oberta de Catalunya (UOC)*

Abstract

University ranking systems are being implemented with the aim of assessing and comparing higher education institutions at a global level. Despite their being increasingly used, rankings are often strongly criticized for their social and economic implications, as well as for limitations in their technical implementation. One of these limitations is that they do not consider the specific characteristics of online education. This study used a participatory approach to define a set of criteria and indicators suitable to reflect the specific nature of distance education. This endeavour will help evaluate and rank online higher education institutions more appropriately than in current practice, where indicators are devised for traditional universities. To this end, several stakeholders and informants were involved in a Delphi study in an attempt to reach the broader higher education institutions (HEI) community. According to the study participants, apart from students' achievements and general quantitative measures of HEI performance, which are quite common in traditional ranking systems, teaching and student learning experience turned out to be the most important criteria. Student support, teacher support, technological infrastructure, research and organization were deemed middle ground criteria, while sustainability and reputation were regarded as the least important criteria.

Keywords: university ranking, online education, quality in higher education, institutional reputation, Delphi study, performance indicators

Introduction

In an increasingly internationalized and globally connected scenario, the higher education arena is becoming an ever more competitive market, with universities under constant pressure to secure student numbers and research funding. As part of this phenomenon, the component of distance education is constantly growing in the framework of higher education, even though some providers of traditional education question its quality.

In such a context, university rankings—such as Academic Ranking of World University (ARWU) and Times Higher Education World University Rankings—have become powerful tools enabling universities, potential students, policy-makers and funders to measure and compare universities at a global level (Brasher, Holmes, & Whitelock, 2017). Despite their increasingly widespread use, university rankings have been strongly criticized for their social and economic implications, as well as for their technical implementation (Amsler & Bolsmann, 2012; Bougnol & Dulà, 2015; Lynch, 2015).

One such limitation is that, despite the crucial role distance higher education and online education providers (such as online universities) are known to play at the European level and worldwide (Li, 2018; High Level Group on the Modernisation of Higher Education, 2013), at present, existing university ranking systems do not consider their specific characteristics (Brasher et al., 2017; King, 2012). This issue has already been recognized by research in the fields of evaluation and assessment of the quality of online education, leading to the development of several benchmarking tools specifically tailored to evaluate the quality of online programmes or courses. However, since those instruments are not designed with the aim of ranking, they cannot be used to compare online higher education institutions (HEIs). There is a risk that current rankings of online universities misrepresent their actual quality, when compared to traditional universities.

It should be noted that the problem of evaluating online education, in general, can be tackled at different levels (from institution level to course level) and in different contexts (pure online versus blended). For example, one can consider the object of the evaluation to be online institutions (such as the open universities), traditional universities running only a few courses, entire programmes through the Internet, or, perhaps, MOOC providers and so on. Defining criteria and indicators for all of these situations (or any possible variant thereof) is an extremely delicate matter, thus each study needs to clearly state its target context. In this study, we focus on evaluation of online HEIs, rather than on individual online courses or programmes.

In order to develop a ranking tool tailored to capture the quality of online HEIs, we first need to understand what criteria and indicators are the most appropriate for measuring the specificities of online universities. In this paper, we present the approach we adopted to address this need and the results obtained, in an attempt to contribute to the debate about how we should valorize online HEIs within existing ranking systems. We seek to address the following research questions:

1. What criteria shall we consider for ranking online HEIs?

2. What indicators (and weights) shall we consider for ranking online HEIs?

To answer these research questions, we took a participatory approach to defining the criteria and indicators (i.e., the definition process involved several stakeholders and informants in an attempt to cover the broader HEI community). This was done on the assumption that taking into consideration the points of view of all relevant informant bodies and individuals is crucial for the criteria and indicators to be understood, recognized, accepted and ultimately used (Usher & Savino, 2006). Furthermore, this approach should lead to a comprehensive set of criteria and indicators that capture and evaluate all the aspects and variants at play.

Our purpose was to identify the features that reflect the peculiarities of distance institutions specifically; thus, we let aside from our study some criteria and indicators which can apply to any HEI (traditional or online), as these can be easily drawn from existing ranking systems, as will be further explained in the following sections.

Setting the Scene

The higher education world is becoming more and more complex, with a growing number of universities and education providers acting as commercial enterprises competing within a global market. According to Leo, Manganello, & Chen (2010), universities have to compete as well as consolidate or improve their reputation. In such a competitive environment, there are rapidly emerging tools that aim to represent the prestige and reputation of universities, generally in terms of perceived quality, by means of qualitative and quantitative indexes. These tools include internal and external quality assurance processes and procedures, accreditation, evaluation, benchmarking, accountability systems and university rankings.

Rankings are an established technique for displaying the comparative position of universities in terms of performance scales. These have become quite popular and are seen as a useful instrument for public information and quality improvement (Vlăsceanu, Grünberg, & Pârlea, 2007). Since 2003, when the ARWU was born, university rankings have been used by HEIs at global level as a means of becoming more visible, reputable and marketable. The most significant university ranking systems include global rankings (e.g., ARWU, Quacquarelli Symonds World University Rankings, Centre for Science and Technology Studies Leiden Ranking, U-Multirank), national rankings (e.g., Centre for Higher Education University Rankings, Guardian, United States News & World Report Best Colleges), and global discipline-focused rankings (e.g., Financial Times Master of Business Administration rankings). All these ranking systems periodically issue lists of ranked universities, based on criteria and indicators that are assessed, measured and then usually aggregated into one (or more) composite measure(s). Criteria and indicators vary from system to system, but they all share the same underpinning philosophy. Typically, the systems rely on self-reported data, provided by the institutions themselves, as well as surveys, bibliometric and patent data, and so on.

According to Bowman and Bastedo (2011), university rankings are said to influence HEIs on different levels, such as institutional aspects (organizational mission, strategy, personnel, recruitment and public relations), reputation, student behaviour, tuition fees and resource supply from external providers. Furthermore,

rankings are also increasingly being used as a policy instrument to assess the performance of institutions by governmental agencies (Salmi & Saroyan, 2007; Sponsler, 2009).

Most of the existing ranking systems are criticized for several methodological shortcomings (Amsler & Bolsmann, 2012; Barron, 2017; Bougnol & Dulà, 2015; Çakır, Acartürk, Alaşehir, & Çilingir, 2015; Lynch, 2015). Among the main weaknesses mentioned, many say rankings are not robust enough, especially as far as validity of indicators, methodological soundness, transparency of sources of information and algorithms, reliability, and so on (Billaut, Bouyssou, & Vincke, 2009; Bonaccorsi & Cicero, 2016; Kroth & Daniel, 2008; Turner, 2013). Furthermore, university ranking systems have been criticized for equity concerns (Cremonini, Westerheijden, Benneworth, & Dauncey, 2014).

Another key limitation of current ranking systems is that the dimension of online education is not represented. That is, most of the criteria and indicators used do not consider online education, so online universities are ranked according to the same indicators used for traditional universities (King, 2012). This shortcoming affects the representation and visibility of online universities, especially in terms of quality; they have quite specific features that set them apart from traditional universities, even though they share many of their goals. Thus, the need to have tools specifically designed to measure and compare the quality of the online services offered by HEIs is emerging with a certain urgency (Kurre, Ladd, Foster, Monahan, & Romano, 2012; Marginson, 2007).

Even if there are no rankings for online universities at present, there are several benchmarking initiatives in the field of online education, emerging from the need to tailor indicators to the specific context of online education. Examples include the Quality Scorecard Suite by the Online Learning Consortium (formerly Sloan Five Pillars), European Foundation for Quality in e-Learning, Quality Matters, and E-xcellence (European Association of Distance Teaching Universities). These tools are designed to assess quality at module and course or programme levels. Therefore, it may prove useful to look at their indicators and guidelines, as these may provide hints about crucial dimensions to focus on for measuring the quality of HEIs. However, since these instruments were not designed with the aim of ranking, they cannot be used to compare online HEIs.

Despite their recognised limitations, university rankings continue to be widely used, especially due to their increasing influence and to the fact that they "satisfy a public demand for transparency and information that institutions and governments have not been able to meet on their own" (Usher & Savino, 2006, p. 38).

Therefore, there is definitely a need to implement a ranking system able to reflect the specific nature of online education, in such a way that online universities are not evaluated by means of inappropriate indicators devised for traditional universities. This, however, presents a number of challenging aspects, not least of which is the need to identify the most suitable criteria and indicators for representing and measuring the nature of online universities.

Context of the Study

This study was undertaken in the context of CODUR Project, a European Erasmus+ project that ended in October 2018. The project aimed to generate a set of quality criteria and indicators for the measurement of

the worldwide online education dimension, and guidelines for integrating these online education quality indicators within other current ranking systems.

In accordance with the above-mentioned objectives, the partnership has dedicated considerable effort to defining a set of criteria and indicators that can be used to rank online HEIs. To this end, the partnership took a participatory approach through a Delphi study, involving informed experts in an effort to cover the broader online HEIs community.

Method

The Delphi method, first proposed by Dalkey and Helmer (1963), is a research technique based on consultation with a panel of experts through multiple questionnaire rounds. In Delphi studies, the results of previous rounds are usually used to prepare questionnaires for subsequent rounds, and participants remain anonymous and work independently.

The CODUR Project's Delphi study remotely involved a worldwide sample of experts, with the aim of defining the criteria and observable indicators for assessing the quality of online HEIs. Moreover, the Delphi study structure also allowed participants to suggest criteria and indicators that were not initially considered. More specifically, the Delphi aimed to:

- identify the criteria relevant for online educational institutions (Round 1);

- determine a non-ambiguous, agreed-upon definition for each criterion (Round 1);

- operationalize each criterion, by identifying multiple observable indicators (Round 2); and

- assign a relative weight to each indicator (Round 2).

The Delphi study research design is represented in Figure 1. In our case, since we started with a solid knowledge base, two rounds were deemed sufficient for the Delphi study (Iqbal & Pipon-Young, 2007), considering that a session of discussion using the Metaplan technique (http://www.metaplan.com/en/) laid the foundation for the first round with the experts (Jones & Hunter, 1995).

```
┌─────────────────────────────────┐
│  Preliminary definition of a set of criteria  │
│  (based on analysis of state of art of online  │
│  education and analysis of existing tools)  │
└─────────────────────────────────┘
              │
              ▼
┌─────────────────────────────────┐
│  Identification and recruitment of experts  │
└─────────────────────────────────┘
              │
              ▼
┌─────────────────────────────────┐
│         Delphi Round 1          │
│         (about criteria)        │
└─────────────────────────────────┘
              │
              ▼
┌─────────────────────────────────┐
│      Round 1 data analysis      │
└─────────────────────────────────┘
              │
              ▼
       ╱─────────────────╲
      ╱   Report about     ╲
     ╱    Round 1 results    ╲
    ╱─────────────────────────╲
              │
              ▼
┌─────────────────────────────────┐
│         Delphi Round 2          │
│   (about indicators and weights)  │
└─────────────────────────────────┘
              │
              ▼
┌─────────────────────────────────┐
│      Round 2 data analysis      │
└─────────────────────────────────┘
              │
              ▼
┌─────────────────────────────────┐
│  Final list of criteria, indicators and weights  │
└─────────────────────────────────┘
```
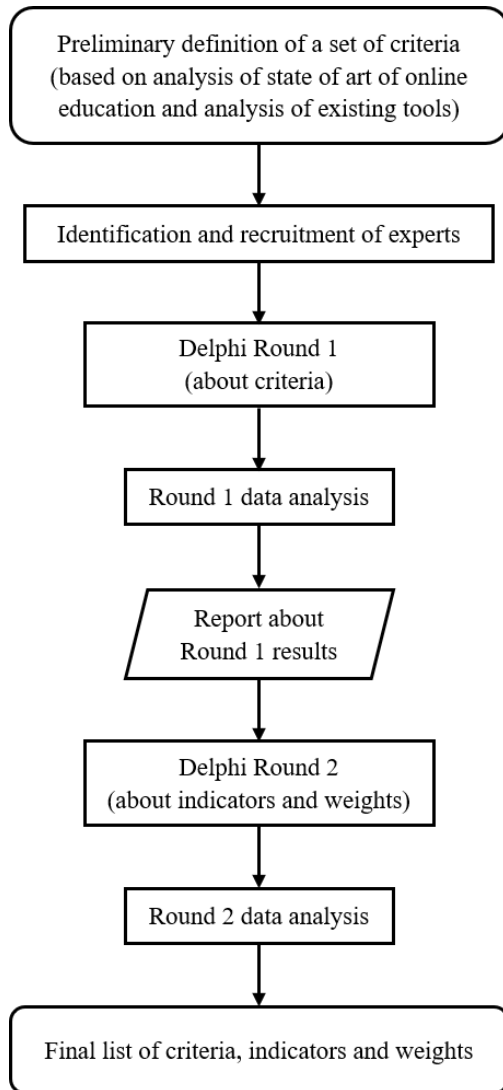
Figure 1. Delphi study research design.

In preparation for Round 1 of the Delphi study, a comprehensive analysis of the state of art in online education across the globe (Giardina, Guitert, & Sangrà, 2017) and a parallel study on the existing ranking systems and benchmarking tools (Brasher et al., 2017) were conducted. Presenting the results of these two analyses is outside the scope of this study. However, it is important to stress that the former analysis pointed out that online education is a global trend, growing at an accelerating rate; this confirmed the need for tools capable of evaluating and comparing the quality of the services offered. At the same time, the latter study, focusing on the indicators used in the existing ranking systems and in some of the most commonly used benchmarking tools, pointed out that we are far from having a unique or standardized way to measure quality and rank universities. On the contrary, the existing systems are heterogeneous, and the indicators adopted are very different and not always clearly defined and/or transparent. As a consequence, the results of these two analyses, which consisted of a huge set of heterogeneous indicators, were not directly usable as input proposal for the Delphi study, but the project partners needed to carry out an intermediate step aimed

at creating a more homogenous and limited set of indicators. To do so, the Metaplan technique was used as a collaborative method for the partners to define a more coherent and smaller set of indicators (Pozzi, Manganello, Passarelli, & Persico, 2017). As represented in Figure 1, the results of this initial phase led to the identification of preliminary criteria and indicators for the quality assessment of online HEIs which was then used as the starting point for the Delphi study. A questionnaire was prepared for Round 1 which, starting from the preliminary set put forward by the project, aimed to collect the experts' opinions regarding the proposed criteria and their definitions.

A panel of experts was recruited (details about the experts and related recruitment follow below) and each one was asked to individually fill in the first questionnaire. After Round 1, the experts were presented with an anonymized summary of the results and a second questionnaire, to trigger Round 2 (see Figure 1).

Both Round 1 and 2 of the Delphi study were carried out using LimeSurvey. The order of presentation of criteria and of indicators within each criterion was fully randomized for each participant.

Participants

The experts involved in the Delphi study consisted of a group of informed stakeholders including (a) researchers on online education; (b) educators working in online HEIs; (c) cross-faculty heads of e-learning; and (d) quality assurance (QA) professionals in traditional, hybrid and online HEIs. The list of experts to be contacted was drawn from the professional networks of the project partners, which include two reputable online HEIs, and great care was taken to include high-profile informants. The list included 140 well-known researchers and teachers—at international level—with expertise in the field of online education and/or HEI policy.

The process of recruiting experts was managed in two steps. A first e-mail was sent to formally invite each expert to participate to the Delphi study; a second e-mail was sent only to the experts who had agreed to participate, in order to provide them with all the information and the link to the questionnaire.

There were 40 participants in Round 1 of the Delphi study. Of these, there were 17 females, 19 males and 4 undisclosed. Participants' ages ranged from 35 to 72 ($M = 53.95$, $SD = 9.47$). Most of the participants (31) were from European countries, while 6 were from Australia, 1 from Israel, 1 from Canada and 1 undisclosed.

Fifteen participants reported being both researchers and educators; 12 were educators, but not researchers, and 8 were researchers, but not educators. The remaining 5 reported being either cross-faculty heads of e-learning or QA professionals in HEIs (see Figure 2). When asked how much they considered themselves informed about University ranking systems, 13 participants reported being slightly informed, 15 stated they were well informed and 11 were very well informed. No participant reported being not at all informed about ranking systems.
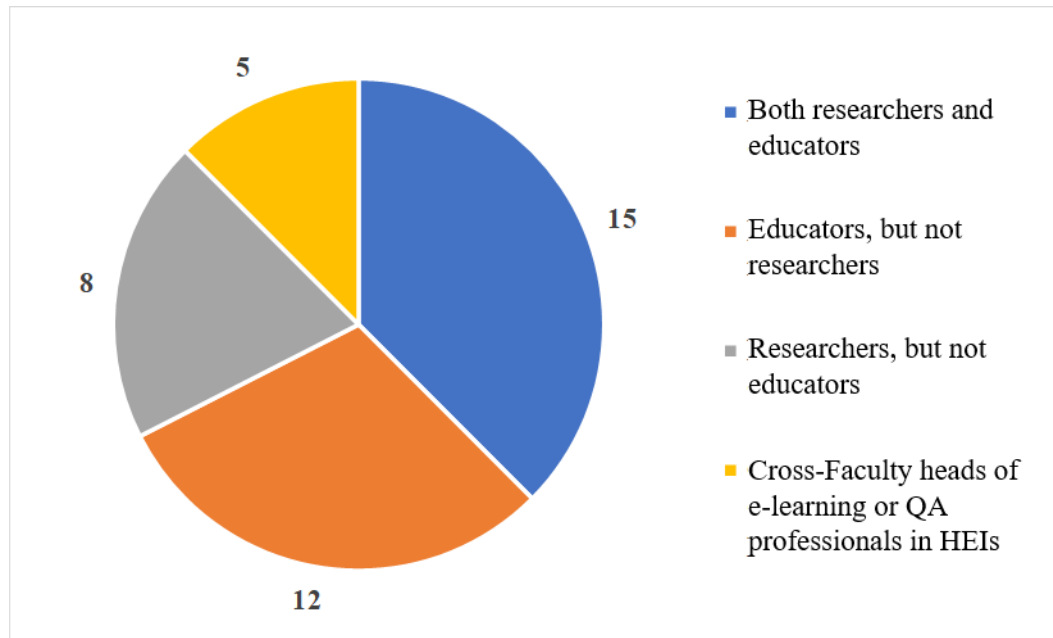
Figure 2. Round 1 Delphi study's experts' background.

All the experts who took part in Round 1 of the Delphi were invited to participate in Round 2. Of the 40 Round 1 participants, 21 took part in Round 2. A retention rate of 46.7% can be considered satisfactory for the scope of this study, and it is aligned with what is reported in literature (Hall, Smith, Heffernan, & Fackrell, 2018).

Results

At the beginning of the Delphi (Round 1) we proposed to the experts a set of nine initial criteria, along with their definitions, for the assessment of quality of online HEIs. The initial proposed criteria, as they came out from our analysis of the literature and existing tools, included:

- quality of learning experience,

- quality of teaching,

- quality of student support,

- quality of teacher support,

- reputation/impact,

- quality of research,

- quality of organization,

- sustainability of the institution, and

- quality of the technological infrastructure.

It should be noted that we intentionally avoided considering the performance of the HEI in terms of student achievement. Such criterion is usually present in the ranking systems we analysed, with indicators such as drop outs, student graduation rate, number of students graduated on time and so on. These indicators are outside the scope of this study, because they are not particular to online HEIs only and should be considered fundamental in any ranking system. Other criteria, such as quality of research, quality of organization, sustainability of the institution, and reputation/impact, which are also in principle applicable to traditional HEIs, it was decided these should be investigated in the study because the very nature of online institutions might change the related indicators.

Participants were asked to rank the criteria in order of importance and provide feedback on the suggested definitions. Moreover, they were asked to give suggestions regarding possible indicators for each criterion.

The resulting rankings were analyzed using Thurstone Case V Scaling (Thurstone, 1927), leading to a relative estimate of importance on an arbitrary scale (Round 1 data analysis). Criteria importance ratings are reported in Figure 3, with associated 95% bootstrapped confidence intervals. Asking participants to rank criteria, instead of judging the importance of each one separately, avoided the possibility that all criteria would be ranked as very important.
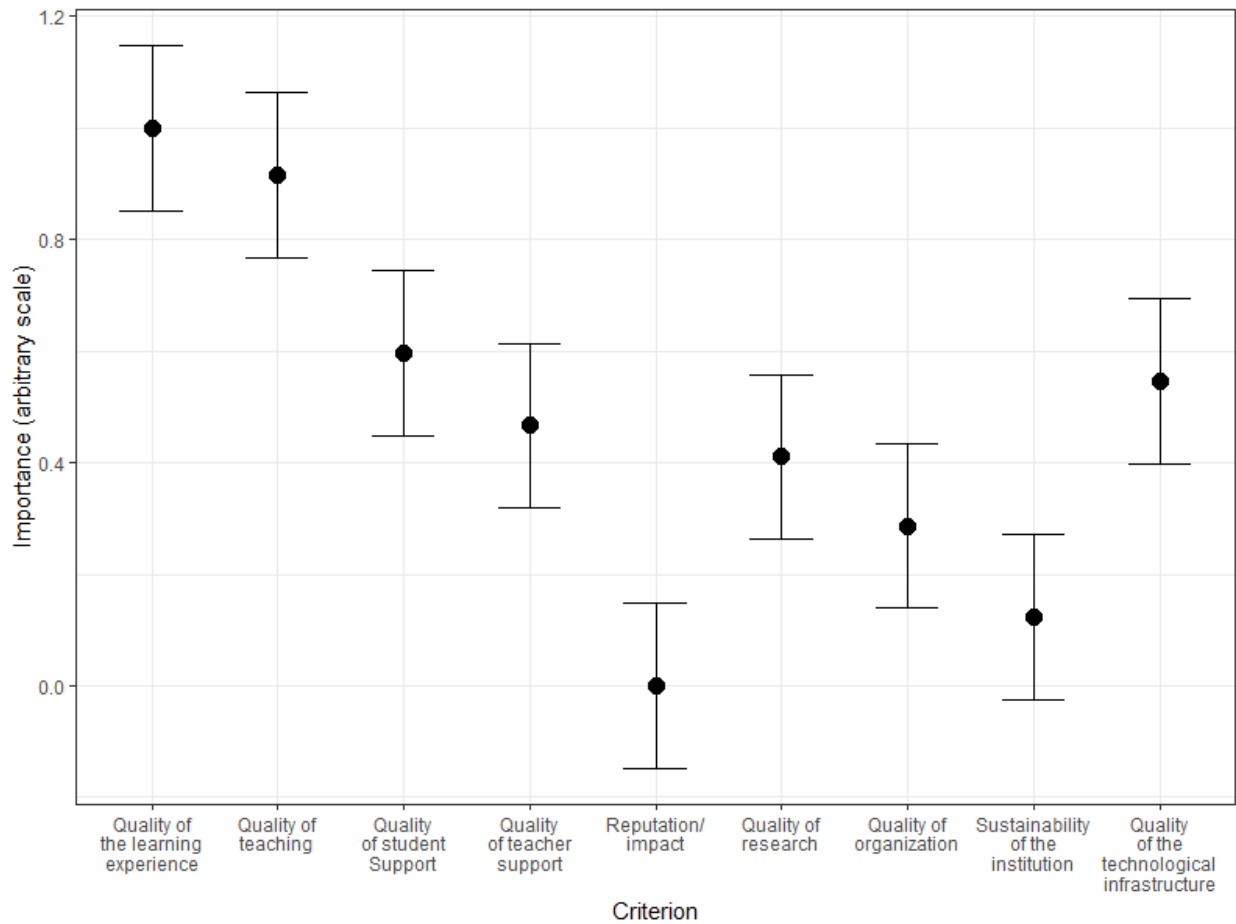
Figure 3. Delphi Round 1 data analysis. Relative importance of criteria with 95% confidence intervals, as estimated by Thurstone Case V Scaling.

According to the results reported in Figure 3, quality of teaching and quality of learning experience were the most important criteria; quality of student support, quality of teacher support, quality of the technological infrastructure, quality of research and quality of organization were deemed middle ground criteria. Sustainability of the institution and reputation/impact were regarded as the least important criteria.

Participants could also suggest adding or deleting criteria. While no deletions were proposed, participants suggested merging the criteria quality of teaching and quality of learning experience, which were perceived as highly overlapping.

This suggestion was accepted (see Table 1), and Round 2 of the Delphi was based on eight—rather than nine—criteria. Minor suggestions on rephrasing criteria definitions were also used to inform Round 2 of the Delphi. Table 1 contains the list of criteria and agreed definitions as they resulted from Round 1.

Table 1

List of Agreed Criteria and Their Definitions (Outcome of Round 1)

Criteria	Definitions
Quality of teaching and learning	The ability of the online HEI to offer effective learning experiences, in terms of sound design, delivery, adopted methods, learning materials, assessment means, and so on.
Quality of student support	The ability of the online HEI to provide support to learners in different areas (e.g., learning, orientation, socializing with peers, organisational issues, and use of technology).
Quality of teacher support	The ability of the online HEI to provide support to teachers and lectures in terms of training provision, organisational issues, use of technology, and so on.
Reputation/Impact	Impact on job market, institutional image, communication strategies, and so on.
Quality of research	The ability of the online HEI to carry out research initiatives and innovation projects.
Quality of organization	Availability of service's structures, efficiency of bureaucracy, and so on.
Sustainability of the institution	Sustainability includes aspects such as the size of the institution, availability of standardised procedures and strategic plans, resources, and so on.
Quality of the technological infrastructure	The ability of the online HEI to offer a sound technological platform, in terms of usability, accessibility, flexibility, types of features offered, and so on.

In preparation for the second round of the Delphi, we addressed the extensive list of observable indicators (75 in total) deemed relevant by Round 1 participants. The goal of the second round was to reduce this list to a manageable number of indicators (roughly half), selecting only those indicators considered most important by the experts we surveyed. We asked participants to choose at most half of the proposed indicators for each criterion. The number of experts who selected each indicator was used both to guide indicator selection and to obtain an estimate of the weight to be applied.

Table 2 reports the final list of 38 indicators chosen by participants as being most important, along with their associated weight. The final indicators we obtained rely on differentiated data sources. These include: (a) student and/or teacher and/or institutional surveys; (b) institutional self-reported data; (c) data coming from the review by an external panel of experts; and (d) bibliometric data.

Table 2

Complete List of Criteria, Observable Indicators, and Weights (Outcome of Round 2)

Criteria	Observable indicators	Weight
Quality of teaching and learning	Student satisfaction of the overall learning experience (through student survey).	16.5%
	Student satisfaction regarding adequacy of the adopted pedagogical approaches to the learning objectives (through student survey).	16.5%
	Institutional support for learning design, in terms of tools, formats, and so on (data provided by the institution).	15.2%
	Percentage of courses that propose personalized paths to reach the learning objectives (e.g., offering different materials or activities depending on culture, learning style, background; data provided by the institution or review by external panel).	11.4%
	Student satisfaction regarding learning materials (through student survey).	15.2%
	Percentage of courses and examinations that make use of diverse forms of assessment (e.g., quantitative and qualitative approaches, human-based and technology-based tools; data provided by the institution or review by external panel).	13.9%
	Student and teacher satisfaction regarding performance reports (through student and teacher survey).	11.4%
Quality of student support	Student satisfaction regarding interactions with teachers and tutors (through student survey).	55.6%
	Student satisfaction with technology support (including helpdesk, FAQ, wizards, support material and initial training; through student survey).	44.4%
Quality of teacher support	Teacher/tutor satisfaction with technology support (including helpdesk, FAQ, wizards, support material and initial training; through teacher survey).	34.7%
	Number of hours of training per year devoted to teaching staff about online learning (data provided by the institution).	22.4%
	Teacher and tutor satisfaction of training opportunities (through teacher survey).	24.5%
	Teacher and tutor satisfaction with feedback on their courses derived from students' surveys (through teacher survey).	18.4%

Criteria	Observable indicators	Weight
Reputation/Impact	Percentage of credits given in service-learning activities, in relation to total number of credits. (Service learning involves students in community service activities and applies the experience to personal and academic development; it takes place outside the HEI; data provided by the institution).	11.0%
	Number of clicks/likes/shares/comments/followers/impressions on academic social networks, such as Academia.edu, ResearchGate, and so on (data provided by the institution).	8.2%
	Percentage of post-graduates actively engaged after graduation (data provided by the institution).	11.0%
	Percentage of former students employed in job sectors matching their degree (data provided by the institution).	19.2%
	A composite measure taking into account the existence of joint/dual degree programmes, the inclusion of study periods abroad, the percentage of international (degree and exchange) students, the percentage of international academic staff (data provided by the institution).	16.4%
	The number of student internships (total per year; data provided by the institution).	8.2%
	The number of student mobility (total per year; data provided by the institution).	13.7%
	The proportion of external research revenues, apart from government or local authority core/recurrent grants, that comes from regional sources (i.e., industry, private organisations, charities; data provided by the institution).	12.3%
Quality of research	Internal budget devoted to research on online learning and teaching per full-time equivalent (FTE) academic staff (data provided by the institution).	16.5%
	Percentage of FTE staff involved in research on online learning and teaching (data provided by the institution).	17.7%
	Yearly average number of publications on online teaching and learning per FTE academic staff (WoS or Scopus publications; data provided by the institution or review by external panel).	17.7%
	Yearly average number of publications with authors from other countries per FTE academic staff (WoS or Scopus publications; data provided by the institution or review by external panel).	10.1%
	Internal budget devoted to disciplinary research per FTE academic staff.	10.1%
	External research income concerning disciplinary projects per FTE academic staff.	11.4%

Criteria	Observable indicators	Weight
	Yearly average number of publications per FTE academic staff (WoS or Scopus publications; data provided by the institution or review by external panel).	16.5%
Quality of organization	Percentage of student complaints or appeals solved or closed (data provided by the institution).	23.8%
	Number of FTEs employed for non-instructional, non-technical support services (e.g., providing assistance for admission, financial issues, registration, enrolment) weighted by student satisfaction for the service (data provided by the institution + student survey).	28.6%
	Student satisfaction with rooms, laboratory and library facilities (through student survey).	23.8%
	Student satisfaction with organization (through student survey).	23.8%
Sustainability of the institution	Availability of an Institutional Strategic Plan for Online Learning (online vision statement, online mission statement, online learning goals and action steps; data provided by the institution).	47.2%
	Percentage of curriculum changes resulting from an assessment of student learning (either formal or informal) within a fiscal year; measures increased flexibility within the curriculum development process to better respond to a rapidly changing world; data provided by the institution).	27.7%
	Percentage of total institutional expenditure dedicated to online programmes (data provided by the institution).	25.0%
Quality of the technological infrastructure	Student satisfaction with the overall learning platform (through student survey).	38.9%
	Measure of compliance with the accessibility guidelines WCAG 2.0 (through technical institutional survey).	36.1%
	Measure of interoperability, for example: (a) with external open sites (e.g., social media, DropBox, Google Drive); (b) between learning management systems; (c) information and teaching/learning materials exchange (e.g., LTI, SCORM); (d) single sign-on access control (data provided by the institution).	25.0%

Discussion

Evaluation, accreditation and ranking are critical aspects within the higher education community. However, even if evaluation, accreditation and ranking are somehow all facets of the same question, it is important not to mix them, as they point to different actions, each one with different aims. In this study, we have decided to focus on the ranking area.

The lack of specific indicators for ranking the quality of online education is seen as an urgent gap the higher education community needs to fill. Currently, the rank attributed to online universities is derived from criteria and indicators that were originally conceived to evaluate traditional HEIs; the risk of misinterpreting the actual quality of distance HEIs is very high, especially when they are compared to traditional universities.

In this paper, we have described the participatory approach we took, in order to fill in this gap, with the final aim of contributing to the debate about how we should valorize online HEs within existing ranking systems. In the following, starting from the data coming from the Delphi, we discuss the two research questions presented at the beginning of the paper.

Research Question One: What Criteria Shall We Consider for Ranking Online HEIs?

As a result of the overall process conducted, we came up with a set of 8 criteria and a total of 38 observable indicators deemed by our experts as able to well represent the peculiarities of distance education institutions (see Table 2). To these, we should add all kinds of indicators typically included in any ranking systems and not peculiar of online institutions, such as student performance indicators.

Regarding the eight criteria which the Delphi participants agreed on, we see that the quality of teaching and learning was considered highly important for evaluating online institutions (see Figure 3). Wächter et al. (2015) recommended focusing particular research effort on "adequate and internationally comparable indicators for the quality of teaching" (p. 78), something which seems to be missing or unsatisfactory in most existing ranking systems. This recommendation is reflected in our results, as the experts felt these criteria were the most important ones for measuring the quality of online HEIs.

Within this quality of teaching and learning criterion, our Delphi experts pointed out the importance of the learning design phase, as this should guarantee that effective pedagogical approaches are adopted and aligned with the learning objectives, and that adequate assessment procedures are put in place (see Table 2). In addition, our Delphi study pointed out the significance of teachers' and students' experience; this is also a crucial point that should be considered in any ranking system, possibly by collecting their opinions through surveys. Moreover, the ability of an online institution to offer personalized learning paths has been considered as deserving valorization, as this represents a distinctive element in respect to traditional universities, where personalization can be harder to achieve.

Quality of teacher support and quality of student support were both considered of medium importance (Figure 3). Regarding quality of teacher support, the experts pointed out the importance of measuring the online institution's ability to provide continuous training opportunities for teachers especially in online

learning, as well as the importance of collecting their opinions on the adequacy of the support being offered (see Table 2).

The quality of the technological infrastructure is important in an online institution and should be measured, according to our experts, in terms of students' satisfaction, compliance with accessibility standards and interoperability.

As far as the quality of research is concerned, this was mainly valorized in terms of publications and budget (see Table 2), while it is the authors' opinion that these should be compounded also by measurements of the level of innovation and impact of research, which are hard to capture if we only look at publications and budget devoted to research.

Organization is measured in terms of efficiency of bureaucracy and adequacy of provided facilities (see Table 2). Reputation/impact, and sustainability of the institution, were positioned in the last places (Figure 3); this might be due to the fact that they are the result of all the other criteria.

The results contained in Figure 3, which represents the relative importance of criteria, can be used as preliminary weights, but in order to use them in practice they should be linearly transformed so that the lowest-ranked criterion (reputation) has a weight higher than 0. Ultimately, this is a matter of choosing how much impact the less important criterion should have on the overall ranking (e.g., should it be dropped? Should it matter at least 1%? Or 2%? Or 5%?). This choice cannot be made just by examining data, but requires careful consideration of the potential practical impact of assigning very different—or very similar—weights to the criteria.

It should be noted that some of the criteria are very much intertwined and the boundaries between them are often blurred. For example, quality of student support, quality of teacher support and quality of teaching and learning probably overlap or are strongly correlated. In addition, some of the interviewed experts raised the objection that the criteria should be independent variables. However, if we accept this position, it may become very difficult to find even two orthogonal criteria.

Research Question Two: What Indicators (and Weights) Shall we Consider for Ranking Online HEIs?

One of the most challenging and often questioned aspects of rankings is their ability to capture and measure the complexity of reality with a reasonable number of indicators. Even U-Multirank, which was an attempt to propose a reasonable number of transparent and easy-to-read indicators, has received recommendations to scale down and simplify its indicators (Wächter et al., 2015). With this in mind, in this study we tried to keep the number of indicators as low as possible, and ended up proposing, on average, four to five indicators per criterion.

Interestingly, the observable indicators proposed by participants for the teaching and learning criterion emphasised the (a) pedagogical approaches adopted, (b) learning design phase, (c) personalization opportunities, and (d) kinds of assessment available. These observable indicators seem particularly reasonable and in contrast with some of the indicators adopted by other existing ranking systems (such as U-Multirank), which focus on indicators of outcome (e.g., the percentages of graduations achieved on time

or the number of academic staff members with doctorates), or on other aspects (e.g., library or laboratory facilities), which deal with organizational and infrastructural aspects. Overall, we think the observable indicators proposed by our Delphi study for the teaching and learning criterion are so significant that we recommend their inclusion in other existing ranking systems addressing online or traditional universities. Similarly, we can identify some criteria for which the indicators suggested by our participants clash with those used in peculiar national contexts. For example, in Italy, sustainability is mainly conceptualized in terms of economic resources alone (Ministry of Education, University and Research, 2016). Among the indicators we selected the greatest weight was given to the availability of strategic plan detailing a vision of the place of online learning in the institution, clearly-stated goals and a plan to attain them—all intangible, but crucial, resources. On the other hand, some of the indicators we suggest are already monitored in some national contexts, and therefore adoption of our proposed list of indicators could be easier for some countries. For example, the UK's National Student Survey already collects several indicators related to student satisfaction, which is present, though in different facets, in several of our proposed criteria. In any case, a phase of adaptation to each national context would be necessary in the interest of optimizing resources for data collection, as well as identifying potential issues related to each country's educational infrastructure. More details on national adaptation of CODUR Project's indicators are presented in Pozzi et al. (2017).

As far as the indicators identified for reputation/impact, quality of research, quality of organization and sustainability of the institution, obviously in principle they can be applied to both online and traditional HEIs, but it is interesting to note that the very nature of online institutions often orients the focus of our indicators differently from those in existing ranking systems. For example, while the internal budget devoted to research is an indicator commonly found in other ranking systems (under the quality of research criterion), in our list of indicators it becomes internal budget devoted to research on online learning and teaching. So, even if the criterion might seem the same as those included in other ranking systems, deeper analysis of its internal indicators highlight that there are peculiarities for online HEIs that should be considered.

By contrast, the indicators for student support, teacher support and technological infrastructure concern aspects of stronger importance for online HEIs than for traditional ones. Nevertheless, we believe these latter dimensions should also be considered nowadays by face-to-face institutions, given that blended approaches are becoming increasingly common and important in traditional contexts as well.

Lastly, the observable indicators have been kept as simple, operative and raw as possible. We have tried to avoid complex or aggregated indicators, to support readability and ease of use for the final user. This does not prevent some being subsequently aggregated, in case one wants to provide a synthetic view of data.

In order to address the issue of transparency of indicators, which is often questioned for existing ranking systems, we have provided the weighting for each indicator, to make explicit the relative importance of each one in relation to the dimension under the lens. It should also be noted that these weightings, too, were assigned based on opinions expressed by the experts in the Delphi study.

Conclusions

Given the lack of ranking systems for distance education, the study presented in this paper aimed to devise a set of criteria and indicators reflecting the specific nature of online HEIs that could be integrated with existing ranking systems. Specifically, this study deals with the idea of defining criteria and indicators to be used for a university ranking system through a participatory approach, based on the iterative contribution of a number of expert stakeholders. This study contributes to the development, testing and refinement of representative quality indicators for online education based on consensus at three different levels: (a) the criteria that should be considered when assessing and ranking online HEIs, (b) the observable indicators that should be adopted for each criterion, and (c) the relative weight that should be applied to each observable indicator. The set of criteria and indicators presented in this paper could be considered both as a stand-alone set for a new ranking system, or, probably more wisely, as a sub-set to be integrated within already existing ranking systems, with the purpose of valorizing the online component of the considered institutions. Integrating our criteria and indicators into an existing ranking system would of course call for an additional step of identifying and deleting possible (partial or total) repetitions. This is something the CODUR Project has already started working on, by exploring in collaboration with U-Multirank the feasibility of integration.

Among the main conclusions of this research is the finding that teaching and student learning experience turned out to be of greater importance than all other criteria. Organization, student and teacher support, research and technological infrastructure were found to be middle ground criteria, while sustainability and reputation were deemed the least important.

The study took a participatory approach to the definition of the criteria and indicators; the design phase was not confined within the project boundaries but involved several stakeholders and informants and the broader higher education community, through the Delphi study, as advocated by Usher and Savino (2006). This led to a comprehensive, easy-to-accept set of criteria and indicators able to capture and evaluate all the aspects at play. Among the limitations of this study, we should note that we run the risk of having excluded stakeholders whose point of view should have been considered; for example, our initial list of proposed indicators included several technical indicators for measuring the quality of the technological infrastructure (e.g., server error rates, average response times). These indicators were among those least selected by our participants, and as such, they were not retained in our final indicators list. However, our participant selection and recruitment may have biased the selection of indicators, for example by not including the professionals charged to run and maintain the technological infrastructure of an online HEI.

Another challenge we faced during the study regarded terminology, which is not always uniquely defined, and can lead to misinterpretation. Providing definitions for terms at the beginning of each survey item was the best solution we found to mitigate the risk of misunderstandings, but we cannot guarantee that all the participants interpreted the definitions we provided in the same way.

Future work should focus on testing the indicators, evaluating their validity as well as the effort required to collect them, as this can become a significant barrier to adoption. Additionally, as argued in the discussion, application of the indicators to specific HEI contexts (e.g., countries) could require their fine-tuning and adaptation, to better reflect characteristics specific to each context.

Acknowledgements

References

Amsler, S. S., & Bolsmann, C. (2012). University ranking as social exclusion. *British Journal of Sociology of Education, 33*(2), 283–301. https://doi.org/10.1080/01425692.2011.649835

Barron, G. R. (2017). The Berlin principles on ranking higher education institutions: Limitations, legitimacy, and value conflict. *Higher Education, 73*(2), 317–333. https://doi.org/10.1007/s10734-016-0022-z

Billaut, J. C., Bouyssou, D., & Vincke, P. (2010). Should you believe in the Shanghai ranking? *Scientometrics, 84*(1), 237–263. https://doi.org/10.1007/s11192-009-0115-x

Bonaccorsi, A., & Cicero, T. (2016). Nondeterministic ranking of university departments. *Journal of Informetrics, 10*(1), 224–237. https://doi.org/10.1016/j.joi.2016.01.007

Bougnol, M. L., & Dulá, J. H. (2015). Technical pitfalls in university rankings. *Higher Education, 69*(5), 859–866. https://doi.org/10.1007/s10734-014-9809-y

Bowman, N. A., & Bastedo, M. N. (2011). Anchoring effects in world university rankings: Exploring biases in reputation scores. *Higher Education, 61*(4), 431–444. https://doi.org/10.1007/s10734-010-9339-1

Brasher, A., Holmes, W., & Whitelock, D. (2017). A means for systemic comparisons of current online education quality assurance tools and systems (CODUR Project Deliverable IO1.A2). Retrieved from http://edulab.uoc.edu/wp-content/uploads/2018/06/CODUR-deliverable-IO1-A2.pdf

Çakır, M. P., Acartürk, C., Alaşehir, O., & Çilingir, C. (2015). A comparative analysis of global and national university ranking systems. *Scientometrics, 103*(3), 813–848. https://doi.org/10.1007/s11192-015-1586-6

Cremonini, L., Westerheijden, D. F., Benneworth, P., & Dauncey, H. (2014). In the shadow of celebrity? World-class university policies and public value in higher education. *Higher Education Policy, 27*(3), 341–361. https://doi.org/10.1057/hep.2013.33

Dalkey, N., & Helmer, O. (1963). An experimental application of the Delphi method to the use of experts. *Management Science, 9*(3), 458–467. Retrieved from http://www.jstor.org/stable/2627117

Giardina, F., Guitert, M, & Sangrà, A. (2017). The state of art of online education (CODUR Project Deliverable IO1.A1). Retrieved from http://edulab.uoc.edu/wp-content/uploads/2018/06/CODUR-deliverable-IO1.A1_Stat-of-the-Art.pdf

Hall, D. A., Smith, H., Heffernan, E., & Fackrell, K. (2018). Recruiting and retaining participants in e-Delphi surveys for core outcome set development: Evaluating the COMiT'ID study. *PloS One, 13*(7). https://doi.org/10.1371/journal.pone.0201378

High Level Group on the Modernisation of Higher Education. (2013). *Report to the European Commission on improving the quality of teaching and learning in Europe's higher education institutions.* Retrieved from
https://www.modip.uoc.gr/sites/default/files/files/modernisation_en.pdf

Iqbal, S., & Pipon-Young, L. (2009). The Delphi method. *The Psychologist, 22*(7), 598–600.

Jones, J., & Hunter, D. (1995). Consensus methods for medical and health services research. *The BMJ, 311*(7001), 376–380. https://doi.org/10.1136/bmj.311.7001.376

King, B. (2012). Distance education and dual-mode universities: An Australian perspective. *Open Learning: The Journal of Open, Distance and e-Learning, 27*(1), 9–22. https://doi.org/10.1080/02680513.2012.640781

Kroth, A., & Daniel, H. D. (2008). International university rankings: A critical review of the methodology. *Zeitschrift fur Erziehungswissenschaft, 11*(4), 542–558.

Kurre, F. L., Ladd, L., Foster, M. F., Monahan, M. J., & Romano, D. (2012). The state of higher education in 2012. *Contemporary Issues in Education Research (Online), 5*(4), 233–256. https://doi.org/10.19030/cier.v5i4

Leo, T., Manganello, F., & Chen, N.-S. (2010). From the learning work to the learning adventure. In A. Szucs, & A.W. Tait (Eds.), *Proceedings of the 19th European Distance and e-Learning Network Annual Conference 2010 (EDEN 2010*; pp. 102-108). European Distance and E-Learning Network.

Li, F. (2018). The Expansion of Higher Education and the Returns of Distance Education in China. *The International Review of Research in Open and Distributed Learning, 19*(4). http://dx.doi.org/10.19173/irrodl.v19i4.2881

Lynch, K. (2015). Control by numbers: New managerialism and ranking in higher education. *Critical Studies in Education, 56*(2), 190–207. https://doi.org/10.1080/17508487.2014.949811

Marginson, S. (2007). The public/private divide in higher education: A global revision. *Higher Education, 53*(3), 307–333. https://doi.org/10.1007/s10734-005-8230-y

Ministry of Education, University and Research. (2016, December 12). Autovalutazione, valutazione, accreditamento iniziale e periodico delle sedi e dei corsi di studio universitari [Ministerial decree on self-evaluation, evaluation, initial and periodic accreditation of Universities]. Ministry of Education, University and Research of Italy. Retrieved from http://attiministeriali.miur.it/anno-2016/dicembre/dm-12122016.aspx

Pozzi, F., Manganello, F. Passarelli, & M., Persico, D. (2017). Develop test and refine representative performance online quality education indicators based on common criteria (CODUR Project

Deliverable IO1.A3). Retrieved from http://edulab.uoc.edu/wp-content/uploads/2018/06/CODUR-deliverable-IO1-A3.pdf

Salmi, J., & Saroyan, A. (2007). League tables as policy instruments. *Higher Education Management and Policy, 19*(2), 1–38. http://dx.doi.org/10.1787/17269822

Sponsler, B. A. (2009). *The role and relevance of rankings in Higher Education policymaking.* Washington, DC: Institute for Higher Education Policy. Retrieved from https://files.eric.ed.gov/fulltext/ED506752.pdf

Thurstone, L. L. (1927). A law of comparative judgment. *Psychological Review, 34*(4), 273–286. http://dx.doi.org/10.1037/h0070288

Turner, D. A. (2013). World class universities and international rankings. *Ethics in Science and Environmental Politics, 13*(2), 167–176. https://doi.org/10.3354/esep00132

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

Vlăsceanu, L., Grünberg, L., & Pârlea, D. (2007). *Quality assurance and accreditation: A glossary of basic terms and definitions.* Bucharest, Romania: UNESCO. Retrieved from http://unesdoc.unesco.org/images/0013/001346/134621e.pdf

Wächter, B., Kelo, M., Lam, Q., Effertz, P., Jost, C., and Kottowski, S. (2015). *University quality indicators: A critical assessment.* Retrieved from http://www.europarl.europa.eu/RegData/etudes/STUD/2015/563377/IPOL_STU%282015%29563377_EN.pdf