

Comparing the comparators: How should the quality of education offered by online universities be evaluated?

Andrew Brasher¹  | Denise Whitelock¹ | Wayne Holmes¹ |
Francesca Pozzi²  | Donatella Persico² | Flavio Manganello²  |
Marcello Passarelli² | Albert Sangrà³ 

¹Institute of Educational Technology, The Open University, Milton Keynes, UK

²Istituto per le Tecnologie Didattiche, Consiglio Nazionale delle Ricerche, Genova, Italy

³Psychology and Education Department, Universitat Oberta de Catalunya, Barcelona, Spain

Correspondence

Andrew Brasher, Institute of Educational Technology, The Open University, Walton Hall, Milton Keynes MK7 6AA, UK.
Email: andrew.brasher@open.ac.uk

Funding information

EU Erasmus+ programme, Grant/Award Number: 2016-ES01-KA203-025182

Abstract

Comparing universities and courses is of interest to a variety of stakeholders including potential students, policy makers, news and media organisations, ranking providers, and universities themselves. There are a range of existing university ranking schemes that provide comparisons (e.g. Academic Ranking of World Universities) but typically these are designed with face-to-face teaching and learning in mind. There is also a growing number of quality assurance tools and approaches aimed at ensuring the quality of online education. This article reports on an analysis of a range of quality assurance systems for online higher education institutions and courses, together with variety of ranking systems targeting conventional universities—with the aim of establishing what they measure, and the metrics used to assess quality. The findings from this analytical study were compared with a meta-analysis of the literature about how students choose their university in order to understand how these instruments influence current student opinion and can be used more proactively by universities themselves. The findings reveal: (1) current ranking systems are of limited value for most potential undergraduate

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2022 The Authors. *European Journal of Education* published by John Wiley & Sons Ltd.

students, particularly with reference to online education, (2) comparison systems that can be of value to students from a variety of different backgrounds are likely to be complex to set up and run, and (3) quality indicators that promote both formative and summative evaluation may be beneficial to both institutions and students.

1 | INTRODUCTION

The higher education arena is becoming an ever more competitive market, with universities under constant pressure to secure increasing numbers of students and research funding. It is in this context that comparing universities and courses is of interest to a variety of stakeholders including potential students and universities themselves. There are a range of existing university ranking schemes that provide comparisons but typically these are designed with face-to-face teaching and learning in mind. There are also a number of quality assurance tools and approaches aimed at ensuring quality in online education that have been developed by a variety of providers (e.g. E-xcellence; Rosewell et al., 2017).

We note that a variety of criticisms of rankings have been made, but they are influential nonetheless since we observe that rankings are important: there are “*clear indications that the rankings are causing changes in higher education policies and institutions*” (Kehm & Erkkilä, 2014, p. 6). However, rankings are flawed, since they are perceived to count what can easily be measured, rather than to measure what counts (Locke et al., 2008). There have also been criticisms of the quality of tools for online education, for example that some of the models are composed of an assemblage of benchmarks which do not build on a comprehensive theoretical approach (Masoumi & Lindström, 2012). Other criticisms of rankings include that they employ a variety of conceptual frameworks and volatile methodologies, yet they ignore institutional diversity and lead to homogenization, and many lack useful metrics for academic quality in terms of learning, teaching and assessment (Dill & Soo, 2005; Kehm, 2014; Vught & Ziegele, 2011). This therefore leaves us with the salient question of whether we can find a way to do better comparisons for online education?

The familiar methods of teaching in higher education (Laurillard, 2002) are used by online universities, but are enacted in different ways to support learning in the online environment. In Table 1 we give some examples of differences in the methods used by online universities for each type of learning activity.

Many sources of data for measuring quality can be relevant to both online and face-to-face universities; for example, student and staff questionnaires, peer and internal review, and publicly available data about institutions. Considering the examples given in Table 1, it is apparent that the relevance and significance of many of the quality indicators used in current ranking systems will differ depending on whether a university is operating online or face-to-face. For example, indicators such as academic staff to student ratio should be interpreted differently depending on whether a university is operating online or face-to-face.

Quality of education is a complex, multi-dimensional concept, and universities may be compared along one or more of its different dimensions, for example as *fit for purpose*, *exceptional* or *transformative* (Schindler et al., 2015). Approaches to measuring transformation such as comparing grades on entry versus grades on completion suffer from the problem of imprecision (e.g., there are only five classes of degree); non-standardised grade boundaries across institutions within and across countries, and the sector as a whole; heterogeneity in educational systems; and variations in which learning outcomes are assessed (Wolf et al., 2015). Other proxy measures such as the employment destination of graduates may be confounded by socio-economic factors such as the economic climate (Hazelkorn, 2014). From an analysis of literature about, and examples of, university ranking systems and quality assurance systems for online education, we have developed a process to develop ways for comparing online

TABLE 1 Learning activities and examples of methods used in face-to-face and online environments

Learning activity	Examples of face-to-face methods	Examples of online methods
Acquisition	Lectures and readings	Online readings, audio or videos
Practice	Exercises and problems	Immediate formative feedback to problem solving attempts provided by e-assessment systems
Discussion	Seminars and tutorials	Group activities are not limited by size of physical spaces, can be run asynchronously with learners all over the world
Discovery (Inquiry)	Field trips and laboratory sessions	Simulations and remote access to laboratory instruments

Source: Table constructed by authors using concepts from Laurillard (2002, p. 81).

universities guided by the assertion that there should be a theoretical basis for the indicators used to evaluate a dimension of quality (Masoumi & Lindström, 2012). The purpose of this process is to clarify if and how quality indicators which are theoretically reliable and valid can be put into practice—and enable the consideration of different data sources and their effect on reliability and validity. The process also considers the needs of stakeholders (including students and prospective students, universities, employers, media organisations, ranking providers, policy makers, universities and their staff), who have a variety of different motivations, interests and involvements in comparisons. Ranking systems and quality assurance systems offer different costs and benefits for each stakeholder group, this is why we have considered both types of metrics.

The article is structured as follows. We start by describing our research questions concerning the ways that online and conventional universities have and can be compared, and the perspectives of stakeholders. Next, we outline the methods used to survey literature and review relevant quality and ranking systems. The following section presents our findings on the nature of quality with respect to online education, and the quality related terminology used in the rest of the article. The basis of our process to develop ways for comparing online universities is described; specifically, by using the example of quality as *transformative* to explore the complexities of identifying reliable and valid indicators of transformation, and of putting them into practice. The final conclusions underscore that given the complexity of measuring transformation, policy changes above the level of individual universities are likely to be necessary if changes to current comparison systems are to be achieved.

2 | RESEARCH QUESTIONS

Our overarching research questions were:

- What characteristics and factors should be taken into account when comparing the quality of education provided by online universities?

- How could perspectives of different stakeholders be accounted for when comparing the quality of education provided by online universities?

Our approach centred on the exploration of two related themes: (1) quality assurance systems for online education and (2) university ranking systems designed for conventional universities with face-to-face instruction. For each theme, we aimed to answer two questions:

- a. What are the defining characteristics of systems within the theme?
- b. What range of factors are considered important for determining quality, and what criteria and indicators could be used to measure each factor?

3 | METHOD

Two methodological phases were adopted. The first phase involved synthesising findings reported in the literature obtained through targeted searches of bibliographic databases, with the intention of generating novel interpretations that do not necessarily exist in any one item of literature (Thomas & Harden, 2008). A targeted search was performed, for each theme.

The second phase involved analysing examples of ranking systems and quality assurance tools in the light of findings from the thematic analysis carried out under phase 1. Overall the process was iterative, as issues and topics found in each of the phases inspired a new look at information within the other phase.

3.1 | Literature searches

The following databases were searched: (a) Education Research Complete, (b) Education Abstracts (H.W. Wilson), (c) British Education Index, (d) Educational Administration Abstracts, (e) Library, Information Science & Technology Abstracts, (f) Communication & Mass Media Complete. These databases were chosen because they include journals that were considered likely to publish articles about educational quality from a variety of stakeholder perspectives. The underlying approach consisted of three steps and was the same for both themes 1 and 2.

1. Define search terms and criteria for inclusion.
2. For each search item returned, decide whether it should be included in the set to be analysed based on the criteria.
3. Carry out a thematic analysis on the set that resulted from step 2 (Braun & Clarke, 2006). This phase involved identifying and naming themes within each item, reusing the theme names, modifying them, and/or adding new themes as each new item was read.

3.1.1 | Quality assurance tools for online education

For the first theme, *quality assurance tools for online education* (theme 1), the initial search was confined to articles written in English and published in academic peer-reviewed journals between January 2004 and July 2018. The start date of 2004 was chosen because the first global university rankings appeared in 2003 (Hazelkorn, 2014), and the search was aimed at returning articles from a variety of theoretical, qualitative and quantitative positions. Specifically, the search was for articles whose abstracts included terms referring to education at university level, and terms referring to quality, and terms referring to online learning or education.

This search produced 85 unique hits. Further screening of the article content resulted in a set of 66 papers, for which the inclusion criteria were: (a) analysis related to online learning (e.g., not blended), (b) analysis related to approaches to quality (papers that mentioned quality but did not discuss it any detail were rejected), and (c) directly related to higher education contexts. 22 additional articles were identified using the snowball method (Greenhalgh & Peacock, 2005).

3.1.2 | University ranking systems

A similar approach was followed for the second theme, *university ranking systems* (theme 2), albeit the search criteria used was intended to find articles related to teaching and learning aspects of university ranking systems. The search criteria initially used was for articles whose abstracts included terms referring to ranking of universities, and terms referring to education, learning or teaching. This produced 62 unique hits from scholarly peer reviewed journals. These were treated in the same way as those for theme 1, including filtering, then snowballing to yield a total of 86 hits.

3.2 | A review of relevant tools

Online searches were carried out using keywords and phrases drawn from meta-analyses to identify relevant tools. Two meta-analyses were used as the basis; a study on quality models in online and open education (Ossiannilsson et al., 2015), and the IREG Observatory's inventory of international rankings (IREG Observatory on Academic Ranking & Excellence, 2017). We chose several well-known and widely used quality assurance systems and rankings that exhibit a range of approaches. The quality assurance tools and systems for online education that we have studied in detail are shown in Table 2.

The university ranking systems that we have studied in detail are shown in Table 3. These were chosen for the following reasons: (1) The *Times Higher Education* (THE) ranking contains indicators related to teaching and learning, (2) *U-Multirank* has been developed with the aim of addressing criticisms of other international rankings (Vught & Ziegele, 2011), (3) *U.S. News Best Online Program* focuses on online education.

TABLE 2 Quality assurance systems for online education that were studied; listed by organisation, name of system and URL

Organisation	Quality assurance tool	URL
ACODE	ACODE Benchmarks for Technology Enhanced Learning	https://www.acode.edu.au/course/view.php?id=5
EADTU	E-xcellence	http://e-xcellencelabel.eadtu.eu/
EADTU	OpenupEd Quality Label	http://www.openuped.eu/quality-label
EFQUEL	ECBCheck	http://www.ecb-check.net/
Online Learning Consortium	Administration of Online Programs & scorecard suite	https://onlinelearningconsortium.org/consult/olc-quality-scorecard-suite/
Quality Matters	Quality Matters	https://www.qualitymatters.org/

Source: Authors.

TABLE 3 Ranking systems studied; listed by organisation providing the system, the name of the system and URL

Organisation	University ranking tool	URL
Times Higher Education	World University Rankings	https://www.timeshighereducation.com/world-university-rankings/
EU funded	U-Multirank	http://www.umultirank.org/
U.S. News	Best online program	https://www.usnews.com/education/online-education

Source: Authors.

4 | FINDINGS

4.1 | Nature of quality

A review of definitions of quality in higher education identified two strategies for defining quality (Schindler et al., 2015). The first is to construct a broad definition that targets one central goal or outcome, such as fulfilling a stated mission or vision. Within the literature that adopted this approach Schindler et al. (2015) identified four broad conceptualisations of quality, i.e., quality as (1) purposeful, (2) transformative, (3) exceptional and (4) accountable. In line with Schindler et al. (2015, p. 5), these broad themes may be summarised as

- Purposeful: quality as fulfilling pre-determined specifications, requirements, or standards.
- Exceptional: quality as something distinctive, exclusive and excellent.
- Transformative: quality as effecting positive change in student learning in the affective, cognitive, and psychomotor domains, and personal and professional potential.
- Accountable: quality as accountability to stakeholders, for the optimal use of resources and the delivery of educational products and services with zero defects.

The second strategy for defining quality identified by Schindler et al. (2015) is one of specifying indicators that reflect desirable attributes of higher education as defined by the organisation behind the quality system. Schindler et al. (2015) reviewed indicators from a variety of sources, including those developed by organisations targeting quality in online education. In so doing they identified four distinct categories of indicators: administrative, student support, instructional, and student performance indicators. Many of the publications and quality assurance models reflected this strategy of specifying particular indicators without including a broad definition of quality (Schindler et al., 2015).

4.2 | Terminology

We have aimed to normalise the different terminology that is used to describe similar concepts across the themes in our research questions. Four interrelated concepts emerged concerning the evaluation of quality.

1. Conceptualisation of quality.
2. Category: an aspect of the conceptualisation of quality. Evaluation of a conceptualisation of quality overall requires evaluation of all categories related to that conceptualisation.
3. Criterion: a standard by which a category is assessed. There may be one or more criterion per category.

4. Indicator: a measure which indicates the extent to which a criterion has been met. There may be one or more indicators per criterion.

Every conceptualisation of quality may be represented by one or more categories. Each category may be evaluated with respect to a variety of criteria, and a number of indicators (typically more than one) show the degree to which the criteria has been met. A schematic illustrating the relationship between these concepts is shown in Figure 1.

For example, a *student support* category can include criteria such as the availability of training, information and ongoing support to aid students in interacting with course websites and other information resources. One indicator could be measures of the availability of technical support for students, e.g., the availability of a 24/7 help desk. We use the term *quality framework* to refer to a particular set of categories, criteria and indicators, and the term *quality system* to refer to how a quality framework is put in practice. A quality system will include mechanisms to gather, analyse and process data, and present resulting measures to stakeholders. All the examples we examined (Tables 1 and 2) are unique quality systems in that they each define their own quality framework, and have their own mechanisms to gather, analyse and process data, and present resulting measures to stakeholders.

4.3 | An approach to comparing quality indicators

In their comparison of five national rankings for comparing universities, Dill and Soo (2005) applied a method suggested by Gormley and Weimer (1999). The method involves consideration of six criteria: the validity of the measures; the comprehensiveness of the measures; the relevance, and the comprehensibility of the information provided to consumers; the functionality of the rankings in motivating improvements in teaching and student learning within organisations; and reasonable preparation costs. Dill and Soo excluded Gormley and Weimer's sixth criteria (reasonable preparation costs) on the basis that they were comparing rankings from a variety of countries with different regulatory requirements on the provision of information necessary to generate performance reports (Dill & Soo, 2005, p. 527). However, we contend that it is valuable to consider preparation costs

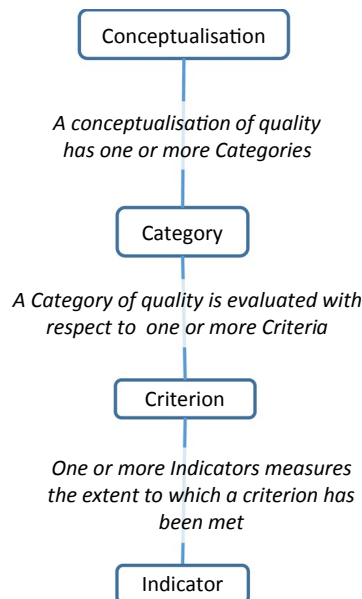


FIGURE 1 A taxonomy of quality. Source: Authors

when comparing methods of measuring quality, because the costs of generating, analysing and representing data necessary will have to be borne by one or more of the stakeholders involved.

Guided by Gormley and Weimer's criteria, we identified three broad sets of characteristics that are of interest.

1. Characteristics of indicator content (i.e., what is measured?). These characteristics relate to what is evaluated and measured by indicators, and the quality factors that these measurements relate to. Comparison of these characteristics centre on Gormley and Weimer's (1999) criteria of validity and comprehensiveness.
2. Characteristics of indicator production (i.e., how is it measured?). These characteristics relate to how indicators are produced. Comparison of these characteristics include consideration of Gormley and Weimer's (1999) criteria of reasonable preparation costs from the points of view of various stakeholders.
3. Characteristics of indicator use (i.e., how can these measurements be used?). These characteristics relate to how indicators are made use of by a variety of stakeholders. Comparison of these characteristics centre on Gormley and Weimer's (1999) criteria of functionality, relevance and comprehensibility from the points of view of various stakeholders.

The nature of what is to be measured will affect how many measures are made, and how the measurement can be used, therefore we consider issues relating to what is measured first. As Masoumi and Lindström (2012) point out, any framework for assuring and enhancing the quality of education should explicitly or implicitly build on a set of theoretical premises. Recalling the four conceptualisations of quality identified by Schindler et al. (2015), we suggest that indicators within any quality framework should relate to one or more conceptualisations of quality, with an explicit link to a theory-based rationale for the selection of indicators. The categorisation of indicators within a quality framework—as measures of purposeful, exceptional, transformative, or accountable quality—will enable potential users of the framework to judge if it is fit for their aims. In this article we focus on *quality as transformation* i.e., the enhancement and empowerment of students, wherein enhancement is understood as a positive change in the knowledge, abilities and skills of students, and empowerment “involves giving power to participants to influence their own transformation” (Harvey & Green, 1993, p. 25). We realise that (1) consideration of a combination of the quality themes will be relevant to many stakeholders, and (2) that the themes are not clearly delineated (i.e., some aspects of quality can be considered to lie within one or more of the themes). Focusing on one theme enables detailed discussion of the issues associated with comparing different quality systems within the limits of a single article, and we examine the limitations of this single-theme approach in our conclusion.

4.3.1 | Characteristics of indicator content

Theme 1: Quality assurance tools for online education

Viewing transformation in terms of educational gain, we adopt the position that “*What best predicts educational gain is measures of educational process: what institutions do with their resources to make the most of whatever students they have*” (Gibbs, 2010, p. 5). To identify possible indicators for quality in terms of transformation, we look to two influential and widely cited models of educational systems. Firstly, Astin's *Input-Environment-Outcome* model (Astin & Antonio, 2012) which resulted from large scale institutional research on the US Higher Education system, through longitudinal surveys of students, teachers and administrators. Responses were collected from millions of students, several hundred thousand college faculty and administrators (Astin & Antonio, 2012). The analysis of the data collected resulted in an empirically derived model in which *inputs* refers to personal qualities that students initially bring to educational programmes; *environment* refers to students' experiences during the educational programme; and *outcomes* refers to skills or talent that educational programmes seek to develop in students (Astin & Antonio, 2012, p. 19).

Secondly, we draw on Biggs' 3P model (Biggs, 1993), which resulted from applying systems theory to education. In this model, Presage variables are those that exist within a university context before a student starts learning and being taught; Process variables are those that characterise what is going on in teaching and learning; and Product variables concern the outcomes of the educational processes. Biggs' 3P model is essentially the same as Astin's *Input-Environment-Outcome* model (Gibbs, 2010, p. 12).

Given that our focus is transformation, and that the best predictor of transformation is the educational environment (or *Process* in Gibbs terminology), we focus our investigation on indicators that can inform us about educational process variables (in Gibbs terms) and the educational environment (in Astin's terms). Astin's empirical foundations backed by Gibbs' theoretical analysis make this three-component model a reasonable choice to inform our comparisons.

Ossiannilsson et al. (2015) reviewed more than forty quality standards, models or guidelines for online learning and produced a table summarising key features of the nineteen most widely used of these. They found that most quality models relate to three main quality categories, each of which is comprised of one or more criterion. The three categories that Ossiannilsson et al. (2015) identified are: (1) *management* (institutional strategy, visions, planning and resourcing); (2) *products* (this includes processes of development, and delivery of curriculum and course modules); and (3) *services* (student, and staff support, information resources etc.).

Schindler et al. (2015) reported one strategy for defining quality as being a matter of specifying indicators that reflect desirable attributes of higher education (Schindler et al., 2015, p. 5). Schindler et al. identified four distinct categories of quality indicators: administrative, student support, instructional, and student performance indicators. These categories can be mapped to those identified by Ossiannilsson et al. (2015) hence supporting the nature of the categorisation that Ossiannilsson et al. (2015) put forward.

We have proposed that comparison of indicator content should centre on Gormley and Weimer's criteria of validity and comprehensiveness, informed by Dill and Soo's (2005) approach to comparing university ranking systems. Dill and Soo discuss Gormley and Weimer's (1999) argument that validity should be evaluated on two dimensions: (1) whether the measures used are clearly linked to valued societal outcomes; (2) whether the measures used control for differences between universities in student characteristics and resources in order to measure the *value added* by a university. The first dimension reflects Locke et al.'s (2008) point about measuring what counts, instead of counting what is measured. The second dimension of *value added* can be applied to any of the conceptions of quality put forward earlier in this article, including the idea of quality as transformative. Dill and Soo (2005) also observe that the validity of measures will depend upon the reliability and accuracy of the relevant data or observations. Consideration of these dependencies is contained in our theme of indicator production, i.e., how is a construct measured?

To compare the comprehensiveness of systems one should consider if and how the range of indicators employed capture the critical dimensions of academic quality. In this paper, we focus on the nature of quality as being synonymous with educational environments that provide transformational learning experiences, wherein quality is characterised as: effecting positive change in students' learning in the affective, cognitive, and psychomotor domains, and personal and professional potential (Schindler et al., 2015). With respect to *comprehensiveness* an approach to comparing quality systems is to consider the categories and criteria identified by Ossiannilsson et al. (2015). By doing this we take Dill and Soo's approach, but extend it to focus on categories and criteria that have been identified as being important to measuring the quality of online education. For each category and criteria, by evaluating: (how) does what is being measured by an indicator relate to theories and knowledge about teaching and learning practices that are known to effect transformational learning experiences?

Theme 2: University ranking systems

A variety of methodologies and data sets are used by organisations producing rankings of universities. However, indicators of educational quality are not present in all systems, and the number and content of the indicators used varies from system to system (Hazelkorn, 2014). When they are present, there is considerable debate over their validity and reliability as true measures of the educational experiences of students, e.g., they "*are often*

those for which data are available rather than being a meaningful measure" (Hazelkorn, 2014, p. 18). For example, staff to student ratio is used as a proxy for educational quality although this ratio will have different implications for various disciplines and learning environments including online education.

As a reaction to the limitations of ranking systems in providing valid measures of educational quality, the Organisation for Economic Co-operation and Development (OECD) ran the Assessing Higher Education Learning Outcomes (AHELO) project (Tremblay et al., 2012). However, weaknesses of this approach have been identified. These include that standardised tests fail to account for the diversity and variation in breadth and depth of discipline-specific knowledge that occurs across different educational contexts (Corbett Broad & Davidson, 2015; Douglass et al., 2012). In addition, a variety of methodological problems have been identified that need to be solved if a standardised testing approach is to be widely used. These are discussed in the next section.

4.3.2 | Characteristics of indicator production

There are three types of processes that need to be carried out to yield a set of indicators that can be used by stakeholders:

1. *Data collection processes*, e.g., peer and internal review, questionnaires (completed by students, staff, and third-parties), publicly available data, Virtual Learning Environment data.
2. *Indicator production processes*, including data analysis, processing and validation).
3. *Indicator representation and publication processes*, including reporting and publication online via interactive, non-interactive sites, copyright issues.

The exact nature of these processes vary according to the aims of the particular system under consideration.

Theme 1: Quality assurance tools for online education

Quality systems for online education such as those described by Ossiannilsson et al. (2015) typically use a process of initial self-evaluation using a scorecard or benchmarks with associated guidance documentation. This may be followed by a peer review to achieve certification. However, consistency of the peer review stage is necessary if ratings are to be used to compare institutions. For example, organisations such as the Online Learning Consortium and Quality Matters provide specifications of online education that can be used by institutions to self-assess the quality of their online courses and programmes. Both organisations run processes for endorsing and certifying the scores that an institution awards itself (Britto et al., 2013). The European Association of Distance Teaching Universities' E-xcellence approach similarly features an internal review followed by a two-day visit by external reviewers (Rosewell et al., 2017). Issues of culture and context affecting the interpretation of the criterion and the scores awarded may arise in this certification process. However, Adair (2017) outlines a procedure for *calibrating* peer reviewers via a process of training, mentoring and certification, aimed at minimising inconsistent reviews. Efforts to achieve consistency and to carry out an external review will involve costs due to workload and travel over and above those required for an internal review.

Theme 2: University ranking systems

In their survey of eighteen university ranking systems Usher and Savino (2007) identified three sources of data on institutions: survey data, university sources, and independent third parties e.g., government agencies. Methods used to gather data about learning and teaching for university ranking systems include surveys of students and reputation surveys (in which invited academics give their opinion about the teaching quality at other universities). Questionnaire based methods have been criticised for a number of reasons including that they are subjective and thus should not be used for ranking (Marginson, 2014).

If we accept that opinions matter, there are still potential problems that arise due to the survey method. In their study of the UK's National Student Survey (NSS), Cheng and Marsh (2010) report that for the years 2005

and 2006 the survey showed very small differences between universities, and more variance in agreement within each university than there is in responses between the different universities. Cheng and Marsh question whether the very small differences between universities are sufficient to help inform the choices of prospective students seeking to compare universities. Pascarella (2001) describes how students' self-reports of the intellectual and academic environment, or self-reported growth in college can be *reasonably reliable and valid* indicators of students' perceptions, but that variations in student characteristics can confound estimates of real institutional impact. With respect to the AHELO method, a variety of methodological issues around the use of tests intended to measure specific skills and competencies across national boundaries have been noted. These include lower student motivation and performance in low-stakes testing conditions (Wolf et al., 2015). With specific reference to transformation, the AHELO approach does not offer production-ready methods that can be used to reliably assess and compare learning gain. Instead, ranking systems typically use proxy indicators that are easy to acquire and satisfy Gormley and Weimer's criteria of reasonable preparation costs.

4.3.3 | Characteristics of indicator use

As noted among others by Çakır et al. (2015); Sandmaung and Ba Khang (2013), there are a variety of actors with interest in the ranking of quality of education.

- Employers
- Media organisations
- Policy makers
- Ranking providers
- Students and prospective students
- Universities and their staff

Moore et al. (2002) make the point that the intended use of any evaluation of distance education generally breaks down into two broad categories: formative and summative. All the stakeholder groups will have an interest in summative evaluation as provided by ranking systems. Universities and their staff are the group that has the greatest interest in formative use of a quality evaluation. However, despite interest in these broad summative and formative categories, "HE [higher education] has multiple purposes with different values placed on these by different stakeholders and not easily sated by the use of single measures" (Evans et al., 2018, p. 8).

Theme 1: Quality assurance tools for online education

Ossiannilsson and colleagues identified four potential uses of quality assurance systems as certification, benchmarking, accreditation, advisory (Ossiannilsson et al., 2015, p. 7). Although any of these functions could potentially relate to one or more conceptions of quality such as transformational education, none of these four functions indicates a commitment to or presence of transformational education per se. Analysis at more detailed levels, i.e., of categories, criteria and indicators, is necessary to determine if and how a specific quality assurance system measures transformation (or any other particular quality factor). From the previous sections in this paper describing indicator content and indicator production it is clear that measures of transformation can be complex, and although awarding of certificates or accreditation can simplify the task of end users such as prospective students, it may mask fine-grained differences which may be of value to students wishing to choose between otherwise similar institutions.

Theme 2: University ranking systems

The approach of traditional ranking systems is to impose a hierarchy upon widely different university offerings which is something the recently developed European-funded U-Multirank system (U-Multirank project, 2017) aims to address. It does this by enabling users to compare universities according to characteristics users themselves

select from a set provided by U-Multirank. However, this makes the process of comparison more complex and time consuming than consulting a ranked list (Barron, 2017).

5 | DISCUSSION

In this article we have reviewed relevant literature and analysed a range of quality assurance systems for online institutions and courses, and a range of ranking systems targeted at conventional universities, to identify ways in which education provided by different online institutions can be compared. We have focused particularly on one quality theme out of four discussed, namely *quality education as transformative*. This enabled us to produce a detailed examination of issues related to this single theme within the limits of this article.

We have identified three aspects that should be considered when deciding on how to compare online universities: (1) what to measure, (2) how to measure the elements of interest, (3) the use of measurements. In the following paragraphs, we discuss how consideration of these aspects with respect to educational transformation illustrate answers to our overarching research questions (a) *What characteristics and factors drawn from themes 1 and 2 should be taken into account?* and (b) *How could perspectives of different stakeholders be accounted for?* Building on Masoumi and Lindström's (2012) assertion that there should be a theoretical basis for the indicators used to evaluate a dimension of quality, the three aspects we have identified can form a process to develop ways of comparing online universities. This process is summarised as a sequence of steps reflecting consideration of what to measure, how to measure it, and how measurements may be used, as shown schematically in Figure 2. In the following paragraphs we discuss issues around each aspect and provide an example to illustrate each step.

5.1 | What to measure

We found that in general ranking systems focus on the outcomes of educational processes, and in so doing provide a summative evaluation of the education provided. Summative approaches to measuring transformation such as

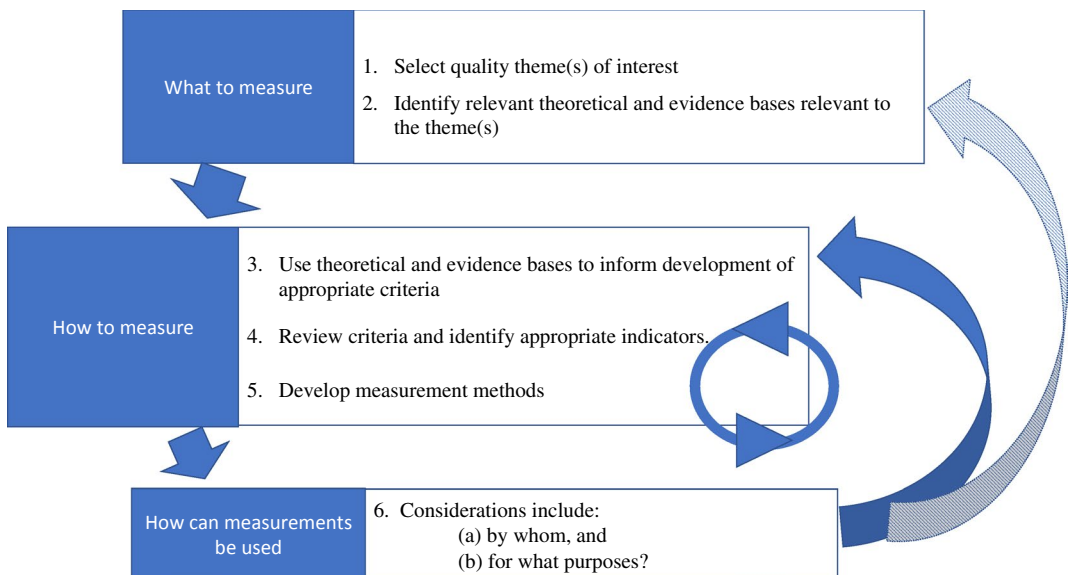


FIGURE 2 The comparing online universities process for developing comparisons of online universities. Source: Authors

comparing grades on entry versus grades on completion suffer from the problem of imprecision (e.g., there are only five classes of degree), non-standardised grade boundaries across institutions within and across countries and the sector as a whole, heterogeneity in educational systems, and variations in which learning outcomes are assessed (Wolf et al., 2015). Other summative proxy measures such as the employment destination of graduates may be confounded by socio-economic factors such as the economic climate (Hazelkorn, 2014).

In contrast, quality assurance systems for online institutions typically provide many indicators of performance of the educational environment (for example for the categories identified by Ossiannilsson et al. (2015) such as curriculum, course design and course delivery). We note that none of the documentation of the quality systems we have examined explicitly mention theories or explicitly described links between educational theories and quality criteria and indicators. However, the documentation of quality assurance systems for online education is generally concerned with enabling institutions to apply an approach to identifying areas for improvement and taking relevant actions. Whilst four of the European tools and a US one mention the term *best practices* in relation to their criteria and indicators, only one (OLC, 2018) includes references to literature in support of the practices described as being *best*. The presence of these references means that it is possible to consult the literature to establish if it includes evidence that supports use of the criteria and indicators. The following provides an example of steps in the Comparing Online Universities Process related to consideration of what to measure.

Step 1: Select quality theme(s) of interest

The first step is to choose one of the four broad conceptualisations of quality that will be the basis of the comparison. This step is necessary because without a specification of what is intended to be measured, there is no way of judging the appropriateness of criteria and indicators to use. For the purposes of this article, we have focused on the theme of quality as transformative: effecting positive change in students' learning in the affective, cognitive, and psychomotor domains, and personal and professional potential (Schindler et al., 2015, p. 5). Although this is a broad definition, it provides a usable basis for identifying criteria and indicators in the following steps.

Step 2: Identify relevant theoretical and evidence bases relevant to the theme(s)

The second step is to identify a relevant theoretical basis, and evidence basis, to use as a means of informing the development and selection of appropriate criteria. For transformation, our literature search identified Biggs' 3P model (Biggs, 1993) as a theoretical basis, and Astin's Input-Environment-Outcome model as the evidence base that supports Biggs' theoretical model. Biggs' theoretical model, supported by Astin's evidence-based model, indicate that the best predictor of transformation is the educational environment, "the student's actual experiences during the educational program" (Astin & Antonio, 2012, p. 19). (The educational environment in Astin's model is equivalent to the *Process* in Biggs' 3P terminology). This focus on the educational environment means that we disregard measures of *Input* or *Outcome* characteristics such as measures of qualities the student brings initially to their university experience, or measures of outcome such as exam results or employment destination. If the focus was on another conceptualisation, such as quality as *Accountability*, consideration of theories from other frames of reference would be necessary, for example theories accounting for the delivery of educational services with low or zero defect rates.

5.2 | How to measure

We now consider the development of criteria appropriate to the focus that was determined in steps 1 and 2. Although ranking systems focus on the outcomes of educational processes, measurement methods used in ranking systems may still be appropriate. The following provides an example of steps for how to measure, from the Comparing Online Universities Process.

Step 3: Using theories and evidence to inform development of appropriate criteria

Step 3 concerns the development of criteria appropriate to the focus that was determined in step 2. In our example on the theme of educational transformation, the focus is on the educational environment. Accordingly,

we have sought to identify theories and evidence relevant to the teaching and learning processes that occur in the educational environment of an online university, to inform the development of criteria. For any conceptualisation of quality there will be a variety of different aspects that combine to represent that conceptualisation. Gibbs (2010) discusses theory and evidence as aspects of process quality and identifies theories and evidence contributing to the overall quality of the educational environment (and hence transformation) including the experience and training of academics, and formative assessment. For our example, we focus on formative assessment. Gikandi et al. (2011) reviewed literature on online formative assessment in higher education and identified several criteria necessary for its effective use. These include that feedback should be timely, ongoing, formatively useful and easy to understand.

Steps 4 concern the identification of appropriate indicators, and step 5 involves development of ways to measure performance with respect to those indicators. These two steps will usually need to be considered together and iteratively, because the nature of the indicators will affect the measurement methods, and the practicality of measurement methods will affect the selection of indicators. Data collected and used by quality assurance systems includes outputs of peer and internal reviews, questionnaires (completed by students, staff, and third parties), and publicly available data.

Step 4: Reviewing criteria and identifying appropriate indicators, and Step 5: Developing measurement methods

Considering the specific criteria for effective online formative assessment identified by Gikandi et al. (2011), indicators could include staff, student and external expert views on the level to which an online university provides formative assessment that is timely, ongoing, formatively useful and easy to understand. Relevant data on these points can be collected by questionnaires targeted at staff and students. The ways that formative assessment is operationalised by an online university includes use of online tools for asynchronous discussion, synchronous conferencing, self-test quizzes, short text answers, and e-portfolios. There is the potential for data to be extracted from these systems and used to create indicators relevant to Gikandi et al.'s criteria.

Marginson (2014) states that rankings should be objective, criticises the use of survey data in ranking systems because surveys deliver data about opinions not the "real world", and raises concerns of the validity of rankings derived from Likert scale data typically used in surveys. The quality assurance frameworks for online institutions we examined all used ordinal scales as measures of their chosen indicators (and hence criteria), and the values are awarded by reviewers from peer institutions. For example, use of the Online Learning Consortium (OLC) framework requires use of the following scale to measure performance against all the various criteria it contains: *Deficient, Developing, Accomplished, Exemplary*.

This peer review process may be called into question on Marginson's objectivity criteria. Issues of culture and context affecting the interpretation of the criterion and the ratings awarded may arise in a process which relies on individual judgement to assign ratings. Inter and intra-rater reliability is an oft-discussed issue (see e.g., Hallgren, 2012). With specific reference to quality assurance of online education Adair (2017) outlines a procedure for *calibrating* peer reviewers via a process of training, mentoring and certification, aimed at minimising inconsistent reviews. Efforts to achieve consistency and to carry out an external review will involve costs due to workload (and potentially travel for external reviewers) over and above those required for an internal review. Adair says that "Peer review is a gift of time, perspective, and expertise from one colleague to another" (Adair, p. 6). However, the reliability of a system based on gifts is questionable. Performing an evaluation of an institution's online education provision will take investment; ensuring that the evaluation is carried out in a way that will be useful beyond the institution itself will require additional investment to ensure inter and intra-rater reliability.

Implementing direct measures of the criteria identified by Gikandi et al. (2011) will be expensive in terms of the number of survey questions needed to generate reliable and valid data, and in addition necessitate more complex processing than a proxy measure. A crucial difference between online universities and conventional universities abilities to provide effective formative assessment lies in the necessity for a reliable technological infrastructure and tools to support assessment and provision of feedback in an online environment. The quality

of online formative assessment and associated feedback depends not only on the qualities of the academic staff involved in developing assessment processes, but also on the nature of the technological system that is developed (or purchased) to carry out these processes. Current ranking systems typically include measures of academic staff's research prowess and some include measures of teaching skills, but only those targeting online universities include measures related to technological infrastructure, for example measures of the variety of educational tools and their reliability (such as mean time between failures).

Analysis of data from a university's online learning environments to quantify aspects of the learning experience of students and teaching experience of teachers has the potential to contribute the kind of objective measures that Marginson recommends. For example in relation to Gikandi et al.'s (2011) *formatively useful* criteria, a systematic measure relating students' engagement with formative feedback to their performance on subsequent learning tasks could be considered. For example, there is evidence that when learners have to wait a long time for feedback, they typically engage with it less once it does arrive (Winstone et al., 2017). Drawing on this, an indicator of timeliness could result from analysis of the relationship between the time that a student completes a learning activity to the time that they download feedback about their performance, with the times systematically recorded by the virtual learning environment that the students engage with. This systematic approach could build on ongoing work in Learning Analytics which although promising has not yet had widespread impact (Ifenthaler & Yau, 2020; Linda et al., 2020). However, examples of operational use of Learning Analytics techniques are emerging as evidenced by Ifenthaler and Yau's review (2020). Clark et al. (2020) identified and validated five critical success factors for the implementation of Learning Analytics at universities, including strategy and policy at an organisational level. If an objective approach to comparing universities using techniques drawing on the Learning Analytics field is to be developed, it will require strategy and policy above the level of individual institutions. Currently, developments in Learning Analytics are supported by institutional, national, or international funding. National or international policy changes would be needed to prioritise research and development of data-analytical approaches to comparison. Policies which are directed at ensuring that data collected and processed for use in comparisons of techniques is useful to individual institutions for formative evaluation of their educational provision are more likely to gain acceptance. Furthermore, the use of survey methods should not be ruled out if they are justified by a theoretical basis. For example, a measure of Gikandi et al.'s (2011) criteria of timeliness could be achieved through surveying students' opinions of the timeliness of the feedback they have received; although this is an opinion, the rationale would be that students are more likely to act on formative assessment that they perceive as timely.

5.3 | Use of measurements

Step 6 involves considering how measurements made by a quality system can be used. There is value in being able to reliably and validly compare the educationally transformative effects of different online universities and courses for all stakeholders we have identified; however, we focus our discussion on the needs of universities and students. Hemsley-Brown and Oplatka (2015) carried out a meta-analysis of literature concerning factors affecting students' choice of university and observed that there is unlikely to be a single list of factors that all students consider; they conclude that the higher education student market is therefore a segmented market. The groups of factors identified by Hemsley-Brown and Oplatka (2015) include demographic and academic characteristics of students. Academic factors such as prior educational achievement constrains the choices available to some students as they will achieve grades which enable them to enrol only to a subset of the university courses available (see the information provided about entry requirements for higher education in the EU, 2018; Hinchey, 2017). Demographic characteristics such as being *time bound* or *place bound* (Pentina & Neeley, 2007) may restrict the options available to prospective students, in some cases to online study only. To be useful to prospective students, a quality system for comparing online universities should reveal facets of the educational experience that

complement those available from the other (institutional, media and social) sources of information available, and hence enhance the decision-making process. A comparison of the transformative potential of universities fits this requirement; however, given the complexity of the data and processing associated with indicators of transformative education it seems likely that it will be difficult to produce a system that will be easy to use for most end users. This is a persisting problem for quality systems. For example, the U-Multirank system has attempted to empower its users by enabling individuals to construct rankings based on factors selected by the individuals. Whilst this approach has been criticised as “costly, time intensive, and defeats the efficiency that gives rankings their purpose” (Barron, 2017, p. 329), Marginson (2014) contends that this approach will improve user satisfaction, though criticises the objectivity of the underlying survey data on which U-Multirank operates. Online universities and online teaching are known to play a crucial role in European Higher Education (McAleese et al., 2014) and around the world (see e.g., Allen & Seaman, 2016); so, despite these criticisms, an approach that yields a detailed and theory-informed comparison of the transformative capabilities of online universities would be of value to prospective students because it would offer a way for prospective students to make more informed decisions. Explicit statements of the theoretical basis for criteria and measures used means limitations of the comparison techniques can be identified. This will be of value to other stakeholders such as policy makers by facilitating judgements of confidence in—and clarification of limitations of—comparisons based on the measures employed.

6 | CONCLUSIONS

This section summarises our findings with respect to our research questions, the first one being: what characteristics and factors should be taken into account when comparing the quality of education provided by online universities? Our review of the characteristics of quality assurance systems for online education and university ranking systems designed for conventional universities with face-to-face instruction showed that ranking systems typically use publicly available data and surveys of institutions as data sources from which their measures are constructed. The indicators used are mainly of the outcomes of the educational process, rather than the process itself. Quality assurance systems for online education provide measures of the quality of the educational process, currently typically through a peer review process. Drawing on literature and examples of both ranking systems and quality assurance systems for online universities, we have put forward a process for developing mechanisms to compare online universities. With respect to our second question—on how perspectives of different stakeholders could be accounted for when comparing the quality of education provided by online universities—we showed how the perspectives of different stakeholders can be taken into account by considering the impact of stakeholder specific choices made at each step in the proposed comparison process.

To demonstrate our comparison method, we have focused on one of four quality themes described; namely, quality education as *transformative*. This choice enabled us to produce a focused examination of issues related to this theme. Current ranking systems are of limited value to most potential undergraduate students because the range of options open to an individual will depend on their prior qualifications, and the way that education is operationalised in online universities is not accounted for. Our Comparing Online Universities Process has indicated that it will be a complex task to set up and run a quality system that enables comparisons of online universities' potential to provide transformative education. For considering one contributory aspect alone (formative assessment), we have demonstrated that a wide range of complex data must be collected and analysed. To put into practice a quality system to compare online universities in terms of transformation, policies above the level of a single institution would be necessary, to ensure that relevant data is available to make comparisons. For prospective students, added value will come from comparison methods that are informed by theory (and so measure what needs to be measured), and which provide information that complements what is available from other sources. For institutions, added value can come from the comparison process providing information that is directly useful to the institution, for example to provide enough detail to enable

institutions to make changes to their offerings. This will depend on the timeliness of the information generated. Quality assurance systems can be set up in a way that private (to the institution) raw information can be used for formative evaluation, in addition to being published in an anonymised and averaged form for use by external agencies for comparison processes. Systems developed following the process we have outlined would contribute valuable evidence informed comparisons for policy makers. Changes such as these should make comparisons useful to institutions, students and policy makers.

We realise that (1) consideration of a combination of the quality themes will be relevant to many stakeholders, and (2) that the themes are not clearly delineated (some aspects of quality can be considered to fall within one or more of the themes). This will necessarily complicate the processes necessary to gather data, analyse and use resulting indicators to produce comparisons. For example, our comparison method includes consideration of *what is measured* (i.e., the characteristics of indicator content), for which we stated that any indicators selected should be linked to a set of theoretical premises. We identified Astin's *Input-Environment-Outcome' model* (Astin & Antonio, 2012) as being a suitable theoretical framework upon which to identify indicators of transformative education. Consideration of other quality themes will add layers of complexity to the comparison process, and the added value of the ensuing complexity will need to be considered on a stakeholder-by-stakeholder basis.

ACKNOWLEDGMENTS

This work was supported by a grant, reference 2016-ES01-KA203-025182, from the EU Erasmus+ programme, 1 October 2016–30 September 2018.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available from the corresponding author upon reasonable request.

ORCID

Andrew Brasher  <https://orcid.org/0000-0002-6672-4699>

Francesca Pozzi  <https://orcid.org/0000-0002-3592-2131>

Flavio Manganello  <https://orcid.org/0000-0001-7584-939X>

Albert Sangrà  <https://orcid.org/0000-0002-5144-9938>

REFERENCES

- Adair, D. (2017). Developing an international pool of certified quality reviewers. *American Journal of Distance Education*, 31(3), 154–159. <https://doi.org/10.1080/08923647.2017.1337875>
- Allen, I. E., & Seaman, J. (2016). *Online report card: Tracking online education in the United States*. Babson Survey Research Group.
- Astin, A. W., & Antonio, A. L. (2012). *Assessment for excellence: The philosophy and practice of assessment and evaluation in higher education*. Rowman & Littlefield.
- Barron, G. R. S. (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>
- Biggs, J. B. (1993). From theory to practice: A cognitive systems approach. *Higher Education Research and Development*, 12(1), 73–85. <https://doi.org/10.1080/0729436930120107>
- Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, 3(2), 77–101. <https://doi.org/10.1191/1478088706qp063oa>
- Britto, M., Ford, C., & Wise, J.-M. (2013). Three institutions, three approaches, one goal: Addressing quality assurance in online learning. *Journal of Asynchronous Learning Networks*, 17(4), 11–23. <https://doi.org/10.24059/olj.v17i4.402>
- 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>

- Cheng, J. H., & Marsh, H. W. (2010). National Student Survey: Are differences between universities and courses reliable and meaningful? *Oxford Review of Education*, 36(6), 693–712. <https://doi.org/10.1080/03054985.2010.491179>
- Clark, J.-A., Liu, Y., & Isaias, P. (2020). Critical success factors for implementing learning analytics in higher education: A mixed-method inquiry. *Australasian Journal of Educational Technology*, 36(6), 89–106. <https://doi.org/10.14742/ajet.6164>
- Corbett Broad, M., & Davidson, P. (2015). *Letter to OECD on AHELO Project*. <https://www.acenet.edu/news-room/Pages/Letter-to-OECD-on-AHELO.aspx>
- Dill, D. D., & Soo, M. (2005). Academic quality, league tables, and public policy: A cross-national analysis of university ranking systems. *Higher Education*, 49(4), 495–533. <https://doi.org/10.1007/s10734-004-1746-8>
- Douglass, J. A., Thomson, G., & Zhao, C.-M. (2012). The learning outcomes race: The value of self-reported gains in large research universities. *Higher Education*, 64(3), 317–335. <https://doi.org/10.1007/s10734-011-9496-x>
- EU. (2018). *Admission and entry to university—Your Europe*. https://europa.eu/youreurope/citizens/education/university/admission-entry-conditions/index_en.htm
- Evans, C., Kandiko Howson, C., & Forsythe, A. (2018). Making sense of learning gain in higher education. *Higher Education Pedagogies*, 3(1), 1–45. <https://doi.org/10.1080/23752696.2018.1508360>
- Gibbs, G. (2010). *Dimensions of quality*. Higher Education Academy.
- Gikandi, J. W., Morrow, D., & Davis, N. E. (2011). Online formative assessment in higher education: A review of the literature. *Computers & Education*, 57(4), 2333–2351. <https://doi.org/10.1016/j.compedu.2011.06.004>
- Gormley, W. T., & Weimer, D. L. (1999). *Organizational report cards*. Harvard University Press. <https://doi.org/10.2307/j.ctv22jnsnr>
- Greenhalgh, T., & Peacock, R. (2005). Effectiveness and efficiency of search methods in systematic reviews of complex evidence: Audit of primary sources. *BMJ*, 331(7524), 1064–1065. <https://doi.org/10.1136/bmj.38636.593461.68>
- Hallgren, K. A. (2012). Computing inter-rater reliability for observational data: An overview and tutorial. *Tutorials in Quantitative Methods for Psychology*, 8(1), 23–34. <https://doi.org/10.20982/tqmp.08.1.p023>
- Harvey, L., & Green, D. (1993). Defining quality. *Assessment & Evaluation in Higher Education*, 18(1), 9–34. <https://doi.org/10.1080/0260293930180102>
- Hazelkorn, E. (2014). Reflections on a decade of global rankings: What we've learned and outstanding issues. *European Journal of Education*, 49(1), 12–28. <https://doi.org/10.1111/ejed.12059>
- Hemsley-Brown, J., & Oplatka, I. (2015). University choice: What do we know, what don't we know and what do we still need to find out? *International Journal of Educational Management*, 29(3), 254–274. <https://doi.org/10.1108/IJEM-10-2013-0150>
- Hinchey, R. (2017). *Applying to university overseas: What to consider*. <https://www.timeshighereducation.com/student/advice/applying-university-overseas-what-consider>
- Ifenthaler, D., & Yau, J.-Y.-K. (2020). Utilising learning analytics to support study success in higher education: A systematic review. *Educational Technology Research and Development*, 68(4), 1961–1990. <https://doi.org/10.1007/s11423-020-09788-z>
- IREG Observatory on Academic Ranking and Excellence. (2017). *IREG inventory on international rankings*. <http://ireg-observatory.org/en/inventory-international-rankings/ranking-profile>
- Kehm, B. M. (2014). Global university rankings—Impacts and unintended side effects. *European Journal of Education*, 49(1), 102–112. <https://doi.org/10.1111/ejed.12064>
- Kehm, B. M., & Erkkilä, T. (2014). Editorial: The ranking game. *European Journal of Education*, 49(1), 3–11. <https://doi.org/10.1111/ejed.12062>
- Laurillard, D. (2002). *Rethinking university teaching, a conversational framework for the effective use of learning technologies* (2nd ed.). Routledge Falmer. <https://doi.org/10.4324/9780203160329>
- Linda, C., Maren, S., & Dragan, G. (2020). Learning analytics: Pathways to impact. *Australasian Journal of Educational Technology*, 36(6), 1–6. <https://doi.org/10.14742/ajet.6853>
- Locke, W., Verbik, L., Richardson, J. T., & King, R. (2008). *Counting what is measured or measuring what counts? League tables and their impact on higher education institutions in England*. Report to HEFCE (Bristol, Higher Education Funding Council for England). http://webarchive.nationalarchives.gov.uk/20100202100434/http://www.hefce.ac.uk/pubs/hefce/2008/08_14/
- Marginson, S. (2014). University rankings and social science. *European Journal of Education*, 49(1), 45–59. <https://doi.org/10.1111/ejed.12061>
- Masoumi, D., & Lindström, B. (2012). Quality in e-learning: A framework for promoting and assuring quality in virtual institutions. *Journal of Computer Assisted Learning*, 28(1), 27–41. <https://doi.org/10.1111/j.1365-2729.2011.00440.x>
- McAleese, M., Bladh, A., Bode, C., Muehlfeit, J., Berger, V., & Petrin, T. (2014). *Report to the European Commission on new modes of learning and teaching in Higher Education*. Publications Office of the European Union.
- Moore, M., Lockee, B., & Burton, J. (2002). Measuring success: Evaluation strategies for distance education. *EDUCAUSE Quarterly*, 25(1), 20–26.
- OLC. (2018). *OLC quality scorecard—Benchmarking tools, checklists, & rubrics for evaluating the quality of online learning programs & courses*. <https://onlinelearningconsortium.org/consult/olc-quality-scorecard-administration-online-programs/>

- Ossiannilsson, E., Williams, K., Camilleri, A. F., & Brown, M. (2015). *Quality models in online and open education around the globe: State of the art and recommendations*. International Council for Open and Distance Education (ICDE).
- Pascarella, E. T. (2001). Identifying excellence in undergraduate education are we even close? *Change: The Magazine of Higher Learning*, 33(3), 18–23. <https://doi.org/10.1080/00091380109601796>
- Pentina, I., & Neeley, C. (2007). Differences in characteristics of online versus traditional students: Implications for target marketing. *Journal of Marketing for Higher Education*, 17(1), 49–65. https://doi.org/10.1300/J050v17n01_05
- Rosewell, J., Kear, K., Williams, K., Rodrigo, C., & Sánchez-Elvira, A. (2017). E-xcellence methodology: Lessons learned over ten years of development and implementation. In G. Ubachs & L. Konings (Eds.), *Proceedings of the online, open and flexible higher education conference: Higher education for the future: Accelerating and strengthening innovation*. The Open University, Milton Keynes. <https://conference.eadtu.eu/download2399>
- Sandmaung, M., & Ba Khang, D. (2013). Quality expectations in Thai higher education institutions: Multiple stakeholder perspectives. *Quality Assurance in Education*, 21(3), 260–281. <https://doi.org/10.1108/QAE-11-2012-0044>
- Schindler, L., Puls-Elvidge, S., Welzant, H., & Crawford, L. (2015). Definitions of quality in higher education: A synthesis of the literature. *Higher Learning Research Communications*, 5(3), 3. <https://doi.org/10.18870/hlrc.v5i3.244>
- Thomas, J., & Harden, A. (2008). Methods for the thematic synthesis of qualitative research in systematic reviews. *BMC Medical Research Methodology*, 8(1), 45. <https://doi.org/10.1186/1471-2288-8-45>
- Tremblay, K., Lalancette, D., & Roseveare, D. (2012). *Assessment of higher education learning outcomes*. Feasibility Study Report 1. <http://www.oecd.org/edu/skills-beyond-school/AHELOFSReportVolume1.pdf>
- U-Multirank project. (2017). U-Multirank | Universities compared. Your way. <http://www.umultirank.org/>
- Usher, A., & Savino, M. (2007). A global survey of university ranking and league tables. *Higher Education in Europe*, 32(1), 5–15. <https://doi.org/10.1080/03797720701618831>
- Vught, F. V., & Ziegele, F. (2011). Design and testing the feasibility of a multi-dimensional global university ranking: Final report *European Community, Europe: Consortium for Higher Education and Research Performance Assessment, CHERPA-Network*.
- Winstone, N. E., Nash, R. A., Parker, M., & Rowntree, J. (2017). Supporting learners' agentic engagement with feedback: A systematic review and a taxonomy of recipience processes. *Educational Psychologist*, 52(1), 17–37. <https://doi.org/10.1080/00461520.2016.1207538>
- Wolf, R., Zahner, D., & Benjamin, R. (2015). Methodological challenges in international comparative post-secondary assessment programs: Lessons learned and the road ahead. *Studies in Higher Education*, 40(3), 471–481. <https://doi.org/10.1080/03075079.2015.1004239>

How to cite this article: Brasher, A., Whitelock, D., Holmes, W., Pozzi, F., Persico, D., Manganello, F., Passarelli, M., & Sangrà, A. (2022). Comparing the comparators: How should the quality of education offered by online universities be evaluated? *European Journal of Education*, 00, 1–19. <https://doi.org/10.1111/ejed.12497>