Co-funded by the
Erasmus+ Programme
of the European Union

codur

Creating an Online Dimension for University Rankings

PROJECT DELIVERABLE: IO1.A2

A means for systemic comparisons of current online education quality assurance tools and systems

Andrew Brasher (OU), Wayne Holmes (OU), Denise Whitelock (OU)

Date: 21st December 2017
Version: 1.2 (final)

Universitat Oberta de Catalunya
The Open University
Consiglio Nazionale delle Ricerche

Contents

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 university ranking systems have become powerful tools. Systems such as the Times Higher Education World University Rankings (Times Higher Education, 2017) and the Academic Ranking of World Universities (ShanghaiRanking Consultancy, 2017) enable universities, potential students, policy makers and funders to measure and compare universities at a global level.

However, university ranking systems have been criticised (Amsler & Bolsmann, 2012; Lynch, 2015). For example, although 'learning experience' and 'student satisfaction' are increasingly becoming important, the main ranking systems still tend to focus on traditional measures such as 'reputation' and 'research'. In addition, they employ a variety of conceptual frameworks and volatile methodologies, yet they ignore institutional diversity and lead to homogenization. In addition, they often have ill-defined target audiences (Vught & Ziegele, 2011, p. 27), or have few winners and many losers. The traditional ranking systems also impose a hierarchy upon widely different university offerings (for example, what does it mean for one university to be ranked at 21 and another to be ranked at 31?), which is something the recently developed European-funded U-Multirank system (U-Multirank project, 2017) aims to address by enabling users to compare universities according to specific characteristics.

Most importantly for this project, although online universities are known to play a crucial role in European Higher Education (McAleese *et al.*, 2014), all of the **existing ranking systems ignore the specific characteristics of online universities**.

Despite sharing many goals with traditional universities, online universities (such as UOC and the Open University) use a variety of teaching approaches and delivery mechanisms, and cater for a more disparate range of students and student needs. The existing ranking systems do not measure and thus inevitably undervalue these unique characteristics (or at least the indicators or metrics that they use are inapplicable), making it especially difficult for online universities to compete effectively in the global higher education marketplace.

This is also true of other online education providers, such as the recently emerging MOOC enterprises and Higher Education Open Educational Resources. However, the rapid growth of these low-cost or free online offerings highlights raises questions about the quality of online provision *between the various online providers*, necessary for prospective students to be able to make properly informed choices, which the existing ranking systems are also unable to address.

Accordingly, the CODUR project's two core objectives are:

a) to produce new criteria and indicators that are applicable to online distance educational offerings, that might contribute to a new online dimension in the U-Multirank international ranking system; and
b) to promote and integrate this new dimension into the U-Multirank ranking system and to the 3 principal pan-European Higher Education networks and quality organizations through a series of high-level meetings and events.

In this deliverable, IO1.A2, we describe the project's first steps towards addressing the first core objective (to produce new criteria and indicators applicable to online education), which

involved devising a process for the systematic comparisons of current online education quality assurance tools and systems. This process was designed to enable the CODUR project to compare quality indicators along a continuum of key themes, which are of central importance to the many different stakeholders who will be involved in using and producing indicators of quality in online education.

In Section 2, we describe our approach to the task. In Section 3, we present a summary of current online education quality assurance tools and ranking indicators in the form of a matrix. In Section 1, we analyse this set of quality assurance tools and ranking indicators, and describe the benefits and problems associated with their production and use. In Section 5, we draw upon the analysis presented in Section 1 in order to develop and describe a process for systemic comparisons of current online education quality assurance tools and systems. Finally, in Section 6, we conclude by summarising issues concerning the comparison of quality assurance tools and systems, and by outlining how this deliverable will contribute to future work in CODUR including deliverable IO1.A3 (led by ITD-CNR).

2 Method

2.1 Introduction

As noted above, in order to address the project's first core objective (to produce new criteria and indicators applicable to online education), we devised a process for the systemic comparisons of current online education quality assurance tools and university ranking systems. Our research method centred upon the exploration of two related themes, each involving a number of key questions, which are then combined in our third theme:

1. Quality assurance tools for online education
    a. What are the defining characteristics of state-of-the-art quality assurance tools for online education?
    b. What range of factors are considered important for determining quality in online education, and what criteria and indicators could be used to measure each factor?

2. University ranking systems
    a. What are defining characteristics of the leading ranking systems designed for face-to-face/conventional universities?
    b. What range of factors are considered important for determining quality in face-to-face/conventional universities, and what criteria and indicators are used to measure each factor?

3. University ranking systems and online education
    a. What quality factors are present in university ranking systems' are not included in the current quality assurance tools for online education, but may be useful in future in such tools?

Our approach to address these themes was desk-based online research, which comprised of:

1. A review of relevant literature (beginning with two reports reviewing relevant quality assurance tools: the Nora project report, produced by the Universitat Oberta de Catalunya (UOC, 2015), and a review of quality models in online and open education (Ossiannilsson, Williams, Camilleri, & Brown, 2015).
2. An in-depth analysis and comparison of specific examples of relevant tools.

Ossiannilsson and colleagues reviewed more than forty quality standards, models or guidelines and produced a table summarising the key features of the nineteen most widely used of these. In the NORA project, UOC reviewed five quality tools, four of which were among the nineteen in the Ossiannilsson table. The tools we studied are summarised in the following sections i.e Section 2.2 "Current state-of the art online quality assurance tools" and Section 2.3 "Leading university ranking systems". Throughout, we aimed to normalise the widely different terminology that is used to describe similar concepts across the themes. For example, henceforward by 'indicators' we will also mean 'metrics'. Most importantly, we use the term 'criterion' to signify a characteristic of online provision that might be quality assured, the term 'category' to signify a group of criteria, and the term 'indicator' to signify a measure that contributes to, or forms, a measurement of a particular criteria.

For example, “Support for students” could be a category, while “Staff support for students” could be a criterion and “Number of Staff forum posts” could be an indicator (NB these examples are deliberately random).

2.2 Current state-of the art online quality assurance tools

We undertook an in-depth study of seven quality assurance tools and systems for online education. The quality assurance tools and systems for online education that we have studied are shown in Table 1. These seven tools included three that were present in both the UOC and Ossiannilsson reviews (the fourth tool that was present in both reviews is no longer in operation), two that were included in only one of the reviews, and two that were not included in either review.

The aim of this selection was three-fold: (i) to ensure a reasonably comprehensive yet manageable coverage, (ii) to widen the scope beyond the two reviews, and (iii) to provide an initial comparison with the U-Multirank tool.

Table 1. Quality assurance tools and systems for online education that were studied.

Organisation	Quality Assurance tool	UOC review	Ossiannilsson review	URL
AQU/UOC	Avaluació d'ensenyaments de formació virtual de l'Agència per a la Qualitat del Sistema Universitari de Catalunya	✓		http://www.aqu.cat/universitats/abanseees/virtual.html#.WR2on261tph
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/
OLC	Administration of Online Programs & scorecard suite	✓	✓	https://onlinelearningconsortium.org/consult/olc-quality-scorecard-suite/
Quality Matters	Quality Matters			https://www.qualitymatters.org/
U.S. News	Best online program			https://www.usnews.com/education/online-education

2.3 Leading university ranking systems

As already mentioned, there have been many criticisms of university ranking systems (Amsler & Bolsmann, 2012; Lynch, 2015; Vught & Ziegele, 2011). We sought to deepen our knowledge of the nature of these systems for two reasons. Firstly, to broaden our understanding of the nature of indicators they include but that are not included in the current quality assurance tools for online education. Secondly to increase our understanding

of how these systems can be used by stakeholders such as potential students and media organisations. We are aware that there is a range of both national and intentional ranking systems (Çakır, Acartürk, Alaşehir, & Çilingir, 2015; UOC, 2015). The aim of our work in this deliverable addresses the project's objective to "produce new indicators and criteria that are applicable to online distance educational offerings, which might contribute to a new online dimension in the U-Multirank international ranking system". In order to build upon (rather than replicate) existing work, and as U-Multirank is an international system, we focused our efforts on studying international ranking systems. However, we also drew on UOC's (2015) and Çakır *et al.*s (2015) work in order to include an overview of the indicators used within national ranking systems, and grey literature such as that produced by the IREG Observatory to contribute to our understanding of 'best practice' and stakeholders' views of both national and international rankings (IREG Observatory, 2006, 2015). The university ranking systems targeted at face-to-face/conventional universities that we have studied are shown in Table 2.

Table 2. Ranking tools targeted at face-to-face/conventional universities that were studied.

Organisation	**University ranking tool**	**URL**
Shanghai Ranking Consultancy	ARWU (Academic Ranking of World Universities)	http://www.shanghairanking.com
Times Higher Education	World University Rankings	https://www.timeshighereducation.com/world-university-rankings/
EU funded	U-Multirank	http://www.umultirank.org/
QS Ltd.	QS University Rankings	https://www.topuniversities.com/university-rankings
Centre for Science and Technology Studies (CWTS)	CWTS Leiden Ranking	http://www.leidenranking.com/

3 A2.1. Matrix of current online education quality assurance tools and U-Multirank ranking indicators

3.1 Introduction

A series of matrices were constructed, which together summarise the state-of-the art quality assurance tools for online education. The aim was to identify a core set of indicators used by the quality assurance tools for online education that could be compared in a systematic way.

A summary of the current online education quality assurance tools studied is presented in Matrix 1. This matrix also includes information drawn from the 'Teaching and learning' area of the U-Multirank tool to facilitate the CODUR project's aim of producing indicators and criteria that add a new dimension to the U-Multirank system.

Matrix 1 makes explicit the following descriptors:

- the organisation responsible for the tool (Organisation);
- the title of the tool (Tool);
- the review in which it was included (if it was included) (Rev.);
- the scope(s) of indicators used within the tool (Scope), where I = Institutional, P = Programme, C = Course;
- the total number of indicators within the tool (N);
- the titles of categories of indicators that are included in each tool, expressed in the words used within the documentation and website related to each tool; we have grouped these categories under headings drawing on the work of Ossiannilsson and colleagues. This grouping has been done to help the comparison but it necessarily masks the details. Categories of indicators that we believe do not fit with the heading identifed by Ossiannilsson and colleagues have been group under the heading 'Other indicators'.

U-Multirank's 34 teaching and learning indicators were not grouped into categories by the providers of the tool. It is not possible to show all of the 34 indicators within Matrix 1 because of the space available. However we have shown which of Ossiannilsson *et al.*'s categories we believe to which *one or more* of U-Multirank's teaching and learning indicators apply to, by presence of the text 'UMR'. In Appendix 1 we present a table showing the relationships we have identified between individual U-Multirank indicators and Ossiannilsson *et al.*'s categories. This shows that 19 of U-Multirank's teaching and learning indicators apply to Ossiannilsson *et al.*'s categories whereas 15 do not. We discuss this in section 4.1.4 'Other'.

A 3 dimensional matrix which reports the indicators in more detail is available online: Online ed qa tools + indicators. This 3D matrix together with the summary presented in Matrix 1 enabled us to build on the work of UOC and Ossiannilsson and colleagues. In particular, the matrices enabled us to check and extend this previous work, towards our goal of identifying a set of indicators by which quality assurance tools for online education may be compared.

					Ossiannilsson *et al.*'s categories of indicators						Other indicators	
					Management	**Services**		**Products**				
Organisation	**Tool**	**Rev.**	**Scope**	**N**	**Strategy and management**	**Staff support**	**Student support**	**Curriculum design**	**Course design**	**Course delivery**		
AQU/UOC	Evaluation of online teaching & learning	UOC	P + C	94	Strategic position of the degree programme	Educational programme			Instruction design	Learning assessment	Results	Internal assessment process
EADTU	E-xcellence	O, UOC	P + C	340	Strategic Management	Staff support	Student support	Curriculum design	Course design	Course delivery		
EADTU	OpenupEd Quality Label	O	I + C	32	Strategic management	Staff support	Student support	Curriculum design	Course design, Course level	Course delivery		
EFQUEL	ECBCheck	O, UOC	P + C	51	Information About and Organization of the programme			Programme/ Course Design	E Media Design; Target group Orientation; Quality of the Content; Evaluation & Review	Technology		
OLC Online Learning Consortium	Administration of online programs	O, UOC	P + C	75	Institutional support; Technology support; Evaluation & assessment	Faculty support	Student support	Course structure	Teaching and learning, Course development/ instructional design	Social and student engagement		
Quality Matters	Quality Matters		C	43			Learner Support		Course Overview Introduction; Learning Objectives (Competencies); Assessment and Measurement; Instructional Materials	Course Technology; Course Activities & Learner Interaction; Accessibility and Usability		
U.S. News	Best online program		I	40			Student services and technology			Student engagement; Faculty credentials and training	Peer reputation	
U-Multirank	U-Multirank	UOC	I + P	34	UMR			UMR	UMR	UMR	UMR	

Matrix 1 Summary of the tools studied

3.2 Characteristics of the Indicators

3.2.1 Introduction

There are a variety of perspectives from which the current state of the art of QA tools and ranking indicators can be discussed. We have identified two broad sets of characteristics that are of interest to the CODUR project:

1. Characteristics of indicator content.
   These characteristics relate to what is evaluated and measured by indicators.
2. Characteristics of indicator production and use.
   These characteristics relate to how indicators are produced and made use of by a variety of stakeholders.

Matrix 1 is the starting point for our investigation into the former. We expand on this in section 3.2.2 'Characteristics of indicator content'.

Moore *et al.* make the point that the intended use of any evaluation of distance education generally breaks down into two broad categories: formative and summative (Moore, Lockee, & Burton, 2002). No matter whether the aim is for formative or summative use, 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
   Questionnaires (completed by students, staff, and third-parties); publicly available data, visits,
2. Indicator production processes (data analysis, processing and validation),
3. Indicator representation and publication processes
   Reporting and publication online via interactive, non-interactive sites, copyright issues.

The exact nature of these processes will vary according to the aims of the particular QA tool under consideration, including its emphasis on formative or summative use, and other factors. We examine this in more detail in section 3.2.3 'Characteristics of indicator production and use'.

3.2.2 Characteristics of indicator content

Ossiannilsson *et al.* found that most quality models relate to three main quality factors, each of which is comprised of one or more components as illustrated by Figure 1.

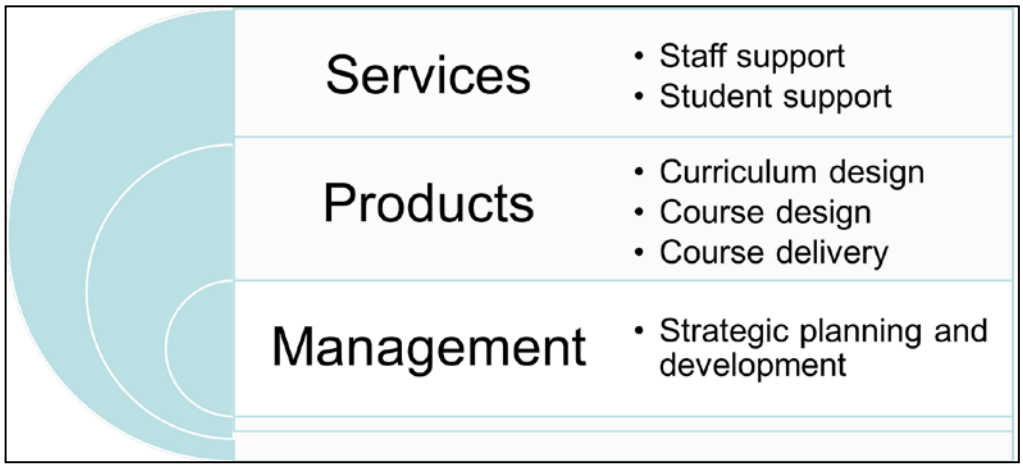

Figure 1 "Three significant main areas related to quality in online learning, including e-learning (Ossiannilsson 2012)" as presented in Ossiannilsson et al. (2015).

This figure is a summary of the three categories that Ossiannilsson *et al.* identified: Management (Institutional strategy, visions, planning and resourcing), Products (including processes of development, and delivery of curriculum and course modules) and Services (student, and staff support, information resources etc.). These factors and their associated sub-categories are the ones that we used to produce Matrix 1. In a particular QA tool, each category may be implemented by a variety of criteria, and associated indicators (typically more than one) that show if and how the criteria has been met or not met.

For example, for the 'Student support' and 'Staff support' sub-categories of the 'Services' category, we note that these can include criteria such as the presence (and quality) of training and informational resources to provide members of staff or students with the skills and background knowledge necessary to engage with online teaching or learning activities. 'Student support' can include criteria such as the presence (and quality) of information provided about programmes such as admission requirements, fees, technological and exam requirements; training, information and ongoing support to aid them in interacting with course websites and other information resources e.g. library sites. 'Staff support' can include criteria such as the presence (and quality) of technical assistance in course development, and availability of training and assistance in online learning, including support to help make the transition from face-to-face to online teaching for those that need it.

Ossiannilsson *et al.* observed variations in the tools they assessed:

> "Differences between the models reviewed lie in the grouping of criteria and the granularity of the detail applied at the performance indicator levels rather than the inherent approach to quality assurance" (Ossiannilsson *et al.*, 2015, p. 7).

Whereas Ossiannilsson *et al.* focused on quality in online teaching and learning, in their study UOC analysed tools for ranking of universities based on university performance across a range of functions. UOC found 4 common quality factors in the tools they studied:

- Teaching and learning
- Research
- Knowledge transfer
- International orientation.

UOC produced a summary of the most common qualitative indicators for each of the factors they identified and for the purposes of CODUR we focus on the 'Teaching and learning' category. In Matrix 2 we propose an alignment between the categories identified by Ossiannilsson *et al.* and the indicators relevant to UOC's 'Teaching and learning' category. Matrix 2 is an edited version of figure 7 from UOC's study: a column identifying the factors from Figure 1 that are relevant to each indicator has been added. One of the tools included in UOC's analysis is no longer in operation, however we have included this tool to show a complete view of the realtionship between the tools at the time the analysis was carried out. The entries relevant to this tool i.e. UNIQe have been presented with a grey background to enable easy identification. We note that UOC also identified the source of data for each indicator as shown in the column headed Src in Matrix 2. We explore potential sources of indicator data in section 3.2.3.

The key for Matrix 2 is as follows.

Key for Matrix 2

Tools	
ECB	ECBCheck
OLC	Administration of online program
UNI	UNIQe European Universities Quality in eLearning (This tool is no longer in operation).
EX	E-XCELLENCE Quality Assessment for e- learning
AQU	Avaluació d'ensenyaments de formació virtual de l'Agència per a la Qualitat del Sistema Universitari de Catalunya.

Category abbreviations	**Categories**
ManIS; ManRes	Management: Institutional strategy & visions; resourcing;
CurricDes; CourseDes; CourseDeliv;	Products (curriculum and course design and delivery)
StudSupp; StaffSup; SuppRes;	Services (student, and staff support, information resources etc.).

(Data) Source	**Src.**
I	Institutional questionnaire
S	Student questionnaire

We note that in the original version of this table the term "Indicator" was used by UOC as the heading of the second column. However, we propose that it would be more apropriate to describe this column as "Criteria" i.e principles or standards by which something may be judged. UOC's document makes it clear that for all of these criteria, there are a variety of diffferent indicators used by each tool to assess whether the criteria have been met.

#	Categories	Criteria	ECB	OLC	UNI	EX	AQU	Src.
1	CourseDes; CourseDeliv	Clear guidelines with information, procedures, timing, tech requirements and other relevant information is available online	X	X	X	X		I/S
2	CourseDes; CourseDeliv	Learning materials and resources are available online or their source is indicated, and are regularly reviewed	X	X	X	X		I
3	CourseDes; CourseDeliv	Student participation, collaboration and interaction (with other students and staff) is encouraged and facilitated online through diverse tools	X	X	X	X	X	I/S
4	CourseDes; CourseDeliv	eTutoring and tech-enabled learner support is available through a variety of means: email, phone, VLE tools...	X	X	X	X	X	I/S
5	StudSupp	Learners' counseling schemes other than academic are available: carrier, psychological...	X	X	X	X		I/S
6	ManIS; ManRes; CourseDes;	Program evaluation procedures are foreseen at the end of the course to evaluate quality of program and contribute to improvement	X	X	X	X	X	I
7	CurricDes; CourseDes; ManIS;	Student feedback is collected and feedback mechanisms in place (regarding program, mentors, tutors...)	X	X	X	X	X	I/S
8	CourseDes; CourseDeliv;	Feedback and assessment of learning outcomes is provided to students, according to pre-set criteria	X	X	X	X	X	I/S
9	StaffSup; SuppRes;	Support related to technology and adoption of elearning tools is available to staff (e.g. Training)		X	X	X	X	I
10	ManIS; ManRes; CourseDes; CourseDeliv;	Design strategy considers usability , navigation and accessibility (considering disabled peoples' needs) and is reliable, secure and effective	X		X	X	X	I
11	ManIS; ManRes; CurricDes; CourseDes;	Flexible learning / institution adapts to learner needs and pace (learner-based approach)	X		X	X		I/S
12	CourseDes; CourseDeliv;	Intended outcomes and methods are clearly stated	X	X	X	X	X	I/S
13	CourseDeliv;	Software is used to detect plagiarism and collusion			X	X		I
14	CurricDes;CourseDes; CourseDeliv;	If blended learning is employed, its rationale is clearly tied to learning outcomes, curriculum content and course strategy	X			X		I
15	ManIS	Professors are asked for feedback*						I

*Matrix 2 Relationships between Ossiannilsson et al.'s categories and indicators from OUOC's 'Teaching and learning' category (*Row 15 is empty apart from the Src. column in the in Nora project report (UOC, 2015))*

In Matrix 2 we note the following frequencies of occurrence of particular factors.

Services

Staff support:	1
Student support	1

Products

Curriculum design	3
Course design	11
Course delivery	9

Management

Resourcing	3
Strategy	5.

3.2.3 Characteristics of indicator production and use

As we mentioned in section 3.2.1, there are 3 processes that need to be carried out to yield a set of indicators that can be used by stakeholders:

1. Data collection process
   Questionnaires (completed by students, staff, and third-parties), publicly available data, visits;
2. Indicator production process (data analysis, processing and validation);
3. Indicator representation and publication
   Reporting and publication online via interactive, non-interactive sites, copyright issues.

In the following paragraphs we put forward a set of factors to describe the processes involved in indicator production and use. For each factor, we identify the group (or groups) of processes that it relates to.

Factor: Scope

Relates to data collection, indicator production, indicator representation and publication.

Nordkvelle, Fossland & Nettleland (2013) reported that concepts of quality can be reviewed at Macro (National/international), Meso (institutional) and Micro (course or module practice) levels (as cited in Ossiannilsson *et al.*, 2015). We will use the term 'Scope' as 'level' is often used to refer to a stage or year within a degree programme. We note that scope is also a factor of indicator content in that an indicator can pertain to e.g. an institution's performance or to a course's performance. However, data may be collected at a different scope to that which the indicator is intending to target (sometimes at a reduced scope due to resource issues).

Factor: Data type

Relates to data collection, indicator production, indicator representation and publication.

The data collected for, and used to represent the value of, the indicators summarised in Matrix 1 fall into a variety of types. Some are categorical, and may be recorded or represented on a nominal, dichotomous or ordinal scale. Examples of ordinal data includes that for indicators for which performance of a particular function is judged to be one of 'Deficient', 'Developing', 'Accomplished' or 'Exemplary'. Other indicators are quantitative, e.g. staff/student ratios.

Factor: Data source

Relates to data collection.

Some data of interest will exist for other purposes (e.g. research publication data), others will be sources which are set up principally for use with a particular QA tool. A few examples are:

Data sources which generate data principally for QA tools

- Bespoke questionnaires. Usually sent to by stakeholders such as staff at the institution, students, peers from other institutions, or expert reviewers.
- Scorecards. Usually filled in by stakeholders such as staff at the institution, or peers from other institutions, or expert reviewers.
- Review documents. Documents produced by staff at the institution, for scrutiny by expert reviewers as part of the review process.

Data sources which exist for other purposes (or independent of QA)

- Statistics about number of teachers and teaching related staff.
- Data about qualifications of staff (including teaching qualifications).

Factor: **Indicator production**

Relates to production process.

Workload: is an indicator produced by manual or machine processing? Or a combination of both?

Ownership: who produces an indicator, or set of indicators? E.g. is it produced by staff of an institution or course, or a third-party such as a media organisation or regulatory authority?

Algorithm: is the mechanism by which a particular indicator is generated from a set of source data described explicitly? E.g. so it can be repeated by others.

Validation: are there steps to ensure that the results generated by a particular process are as expected?

Factor: **Indicator representation and publication**

Relates to indicator representation and publication.

- In what forms can a particular indicator be represented?
- Is a representation comprehensible to the stakeholders it is intended to serve?
- Is a representations useful to the stakeholders it is intended to serve?

With respect to usefulness, Ossiannilsson *et al.* identified four potential uses of QA tools. Each of these is relevant to one or more of the stakeholder groups:

1. Certification
   "Certification/Label is interpreted as a level of recognition granted by the body originating the quality model, award of the certificate/label will follow some form of review and may be accompanied by a requirement that the reviewed institution commits to an improvement plan and later renewal of certification. The originating bodies have various statuses ranging from semi-formal interest groups to international representative bodies".
2. Benchmarking
   "Benchmarking is a process of comparison of institutional performance with that of others, allocation to the benchmarking group indicates that either the originating organisation operates a benchmarking service or there is evidence of the model having been used in benchmarking exercises".
3. Accreditation
   "Accreditation is interpreted as a form of mandatory certification or licensing of institutions and/or their programmes that grants access to national financial support or recognition of awards for employment purposes. Accreditation is a process operated by formal agencies, such as Ministries of Education, Quality Assurance Agencies and Professional Bodies".
4. Advisory
   "Some of the documents reviewed are designed to solely fulfil advisory purposes offering structured guidance to the issues associated with open, distance and online education but not presenting processes of evaluation or performance measurement"
   (Ossiannilsson *et al.*, 2015, p. 7).

3.3 Matrix relating indicators to quality factors

3.3.1 Presage, process and product

In this section we make use of the work of Gibbs (Gibbs, 2010) in which he identified a range of dimensions of quality and examined the extent to which each could be considered a valid indicator, with reference to the available research evidence. Gibbs' report is not aimed at online education (the report is about quality of education in general) but it does include references to it, and the findings are in general relevant to both face-to-face and online education e.g. the underlying 3Ps model has been used in relation to quality in MOOCs (Hood & Littlejohn, 2016).

Gibbs adopted this commonly used '3P' model (Biggs, 1993), which approaches education as a complex system with 'Presage', 'Process' and 'Product' variables interacting with each other. Note that Biggs (and Gibbs) use the term 'Product' to refer to a different concept to that which Ossiannilsson *et al.* use the same term.

> "Process variables are those that characterise what is going on in teaching and learning, for example class size, the amount of class contact and the extent of feedback to students.

> "Product variables concern the outcomes of the educational processes and include student performance, retention and employability. Products can also include psychometric measures of generic outcomes of higher education, such as students' ability to solve problems. In some studies the key product measure is not student performance, but educational gain: the difference between performance on a particular measure before and after the student's experience of higher education.
>
> "Presage variables are those that exist within a university context before a student starts learning and being taught, and include resources, the degree of student selectivity, the quality of the students, the quality of the academic staff and the nature of the research enterprise. None of these presage variables determine directly how the educational process may be conducted, although they often frame, enable or constrain the form education takes."
> (Gibbs, 2010)

So, whereas Ossiannilsson *et al.* use 'product' to refer to things that are used *within* an educational process, Biggs and Gibbs uses 'product' to mean a *result* of an educational process. In the rest of this report, where there might be some doubt we will make it clear which of these meanings are referring to.

Gibbs observes that the categorisation of variables as presage, process or product is not always straightforward. For example, some process variables such as the level of student engagement may be related to other process variables, such class size, which may in turn be related to funding levels.

We have used both Gibbs and Ossiannilsson *et al.*'s categories to classify ciriteria and indicators from the tools for quality assurance of online education that we have studied. This classification enables us to sort the criteria and indicators by both Ossiannilsson *et al.*'s categories and Gibbs/Biggs categories. We have produced a spreadsheet that enables the user to sort the data according to their interests, for example enabling them to see all the 'Process' citeria and indicators that are relevant to e.g. 'Course Design' in one continous set of rows drawn from all the tools. As there are a total of 643 indicators across these tools it is impractical to show or use this classification on paper so we have made the full matrix available online). However we present a small excerpt from it in Matrix 3 within this document. This shows a few example rows from three of tools, including the criteria, indicator and Gibbs 'P' and Ossiannilsson *et al.* categorisations we have applied. In the online matrix, indicators which do not correspond to Gibbs' and Ossiannilsson *et al.*'s categories are denoted by an entry of 'Other' in the 'P' or Ossiannilsson column. These entries correspond to the entries in the 'Other indicators' columns of Matrix 1.

Tool	Criteria	Sub-criteria	Indicator	P	Ossiannilsson *et al.* (Primary descriptor)	Ossiannilsson *et al.* (Secondary descriptor)
OLC	Learning effectiveness	Teaching and learning	Instructors use specific strategies to create a presence in the course.	Process	Course delivery	Course design
OLC	Student satisfaction	Social and student engagement	Students should be provided a way to interact with other students in an online community (outside the course).	Process	Course delivery	Course design
OLC	Access	Technology support	The technology delivery systems are highly reliable and operable with measurable standards being utilized such as system downtime tracking or task benchmarking.	Presage	Course delivery	
ECB Check	Information About and Organization of the programme	Organizational and technical requirements	Technical requirements necessary to allow for adequate participation in the programme/course are clearly described.	Process	Student support	
US News Best online prog.	Student engagement	Best practices	Use of collaborative coursework	Process	Course design	Curriculum design
US New Best online prog.	Student engagement	Best practices	System to track students after graduation	Product	Management	

Matrix 3 Excerpt from our classification of the tools

4 A2.2. Comparative analysis of quality assurance tools for online education

In this section we compare QA tools within the set outlined in section 3 “A2.1. Matrix of current online education quality assurance tools and U-Multirank ranking indicators”, raise issues and describe the benefits and problems with respect to their production and use with respect to online education.

In section 3.2 we identified a range of characteristics through which comparisons between quality assurance tools can be made. In order to make useful comparisons between QA tools it is necessary to consider the needs and aims of a variety of potential users of the tools. Moore *et al.* make the point that the intended use of any evaluation of distance education generally breaks down into two broad categories: formative and summative (Moore *et al.*, 2002).

There are a variety of actors with interest in the uses of quality assurance tools for online education. Çakır *et al.* (2015) identified four groups with interests in summative tools

- Prospective students and their families
- University administrators
- Policy makers
- Media organisations

(Çakır *et al.*, 2015).

Sandmaung & Ba Khang studied 8 papers on QA tools in HE and identified

- Students
- University teaching staff
- University managerial staff
- Employers

as stakeholders in addition to those groups put forward by Çakır *et al.* (Sandmaung & Ba Khang, 2013).

Both formative and summative use of quality indicators may be of interest to any of the stakeholder groups identified in the two lists above. For example, an individual student or prospective student may be interested in a rankings based on summative indicators in order to compare their (potential) educational experience with others available. Other students may be interested in formative evaluations to inform their choice of next course, e.g. if an institution (or programme or course) that they are interested in announces a plan for change in response to a formative evaluation.

In the following sections we analyse and compare indicators drawn from the set of tools summarised in section 3.1. To do this we use the categories and factors described in section 3.2 ‘Characteristics of the Indicators’ to structure our analysis of the characteristics of indicator content, production and use. In doing so, we bear in mind the usual concerns about any evaluation: i.e. concerns of validity (do the indicators measure what they claim to measure?) and reliability (do the indicators produce the same result in all circumstances under which the same result should be produced?) We also reflect on the implications for each of the stakeholder groups listed above.

4.1 Indicator content

There are approaches to teaching and learning that have been identified as 'best practices' in conventional undergraduate education.

> "The 'Seven Principles of Good Practice in Undergraduate Education' (Chickering and Gamson 1987a, 1987b, 1991) are based on a very wide review of empirical evidence, and have been used widely in the US and elsewhere as guides to the improvement of university teaching." (Gibbs, 2010).

The principles are that good practice:

- encourages student-faculty contact,
- encourages cooperation among students;
- encourages active learning;
- provides prompt feedback;
- emphasizes time on task;
- communicates high expectations; and
- respects diverse talents and ways of learning.

Though these principles are drawn from evidence on conventional face-to-face higher education, Bangert describes them as

> "well suited for guiding the design and delivery of quality Internet-based instruction (Billings, 2000; Graham, Caglitay, Lim, Craner, & Duffy, 2001)"

Other researchers have focused on best practices in relation to specific aspects of online teaching and learning. For example online communities (Kear, 2011), MOOCS (Warburton & Mor, 2015) and serious games (Tsekleves, Cosmas, & Aggoun, 2014).

With all these best practice guides in mind, a general approach to reviewing the suitability of the content of an indicator is to check if it aligns with an aspect of evidence-based best practice. (Some of the tools shown in Matrix 1 claim to be informed by best-practice guidelines with the exception of U-Multirank, i.e. the documentation of E-xcellence, ECB-Check , OpenupEd, OLC , Quality Matters, explicitly mention the term 'best practice' in relation to the elearning). The type of data used to represent the indicator needs to be considered. For example, can an indicator which uses an ordinal scale of excellence be utilised to make valid comparisons between institutions? (E.g. the scale 'Excellent in all aspects', 'Adequate with some examples of excellent performance', 'Adequate', 'Not adequate in some aspects', 'Not adequate in majority of aspects'.)

As we mentioned in Section 2.1, in this document we use the term 'criterion' to signify a characteristic of online provision that might be quality assured, the term 'category' to signify a group of criteria, and the term 'indicator' to signify a measure that contributes to, or forms, a measurement of a particular criteria. However, through our study of the tools listed in Table 1 and Table 2, it is apparent that there is a variety of different terminology used to describe the characteristics of quality indicators. Some tools use the term 'indicator' when 'criterion' might be more appropriate; some use the term 'benchmark' or 'standard' instead of 'criterion'; in some tools, specific indicators for a particular criterion may not be explicitly stated, but a descriptive text suggesting several related relevant facets of behaviour or resources is given. We describe specific examples of the variations between the tools in the remainder of section 1 of this document.

4.1.1 Services

Ossiannilsson *et al.* identified two groups that may need support when engaging in online teaching and learning, i.e. staff and students. The Services category relates to services that are necessary to enable staff and students to engage in online learning, and which can increase participation and engagement by these groups. The support needed includes training and informational resources to provide members of staff or students with the skills and background knowledge necessary to engage with online teaching or learning activities. 'Student support' includes information provided about programmes such as admission requirements, fees, technological and exam requirements; training, information and ongoing support to aid them in interacting with course websites and other information resources e.g. library sites. 'Staff support' includes technical assistance in course development, and availability of training and assistance in online learning, including support to help make the transition from face-to-face to online teaching for those that need it.

4.1.1.1 Student support services

We can see from Matrix 1 that five of the seven online QA tools we analysed included criteria and indicators related to student support, and that U-Multirank does not. Four of the five tools that include Student Support criteria utilise ordinal ranking scales as indicators of how well specific criteria have been met. For example, the OLC tool includes the criterion

> "Throughout the duration of the course/program, students have access to appropriate technical assistance and technical support staff."

This is then explained in more detail, with an example and recommendations:

> "Learning to use online library databases for the first time without expert help can be intimidating. And even if students successfully locate appropriate resources, finding full-text versions of journal articles or arranging for e-document delivery can often be a challenge. Library professionals should provide training and educational opportunities for students and faculty, and may want to view teaching information literacy as part of their job. Instructors should make use of the expert help that library professionals can provide, rather than trying to tackle it themselves, or worse yet, leaving students in the dark.
>
> "**Recommendations**
>
> - Provide access to information on accessing library databases (i.e. a link to library resources, an online tutorial, or at a minimum, the librarian's contact information).
> - Provide services that help students locate relevant information such as a self-paced tutorial or library orientation module.
> - Designate a librarian for online learning and provide students with the designated librarian's contact information, so students have a primary point of contact for questions and troubleshooting"
>
> (Shelton, Saltsman, Holstrom, & Pedersen, 2014, p. 71).

Performance against this criterion is judged on ordinal scale as one of: 'Deficient', 'Developing', 'Accomplished' or 'Exemplary'". It is clear that the user of the tool will have to interpret the descriptive text in reference to the culture and context of the course or programme being evaluated, and bring to mind some indicators relevant to that context. For example, the recommendations could be interpreted as indicators, and the user may feel that if all are present then the performance may be judged to be 'Accomplished' or

'Exemplary'. The OLC tool can also be used in a formative way by institutions to self-assess the performance of their online course and programmes. However, the OLC runs a process for endorsing and certifying the scores that an institution awards itself. This process includes a phase of peer review, so issues of culture and context affecting the interpretation of the criterion and the scores awarded may also arise in this certification process.

The E-xcellence tool offers a similar route of self-assessment followed by optional expert review leading to certification. In the first phase a programme will assess itself using E-xcellence's "Quick Scan online questionnaire". This is intended to give the institution

> "a first orientation on the strengths of your e-learning performance and the potential for improvement" (Annex_2_E-xcellence_QS_2016.pdf).

This "Quick Scan" questionnaire requires users to rate their programme's adherence to each of E-xcellence's 35 criteria on a five point ordinal scale:

- Excellent in all aspects;
- Adequate with some examples of excellent performance;
- Adequate;
- Not adequate in some aspects;
- Not adequate in majority of aspects.

This self-assessment can be the basis for a second phase in which reviewers visit the university and do an on-site assessment. To receive certification (the E-xcellence Associations Label) the institution must integrate E-xcellence criteria into its internal QA-system, thus guaranteeing continuous and repeated use of the E-xcellence benchmarks.

For example, the E-xcellence tool has 5 criteria within its 'Student support' category:

- Students are provided with clear and up-to-date information about their courses, including learning and assessment methods.
- Students are provided with guidelines stating their rights, roles and responsibilities and those of their institution. Guidelines of specific relevance to e-learning include provision of hardware, information on accessibility and expected participation in collaborative activities.
- Social media opportunities are provided in order to build and support student communities. This may be achieved using the institution's VLE or through external social media, as appropriate.
- Students have access to support services including technical helpdesk, administrative support and course choice advice.
- Students have access to learning resources, including online library access, study skills development and a study advisor, and they receive guidelines and training in using these resources.

(The criteria in the other categories are shown in the E-xcellence tab of the online matrix: Online ed qa tools + indicators and the online version of Matrix 3. Note: EADTU use the term 'benchmark' for what we refer to as 'criterion'.)

E-xcellence has 91 indicators within this 'Student support' category. All, or a relevant subset of these indicators can be used for self-assessment or by the expert reviewers. However, it is critical to note that the complexity of the analysis required, though likely to be useful for formative purposes, may make consistency of summative reviews across different institutions or programmes or course difficult to achieve, unless there is some sort of quality control of the reviewing process in place. To further explain the detail of the E-xcellence

tool, the 91 indicators within the 'Student support' category are arranged under 6 sub-categories, each of which has 2 to 4 sub-sub-categories. For example, one of the sub-categories is "Pedagogical support" which itself has 2 sub-sub-categories i.e. "Advice and guidance on study skills development" and "Support for e-learning skills development". The "Support for e-learning skills development" sub-sub category has 8 indicators such as "Well-designed online guides/ webpages/ video-tutorials for the IT tools required for students' e-learning studies (virtual campus, software, virtual tools, etc.) are available" and "New students are offered specific online support for the development of required skills and competences for e-learning" (EADTU, 2016, p. 143).

The E-xcellence documentation states that not all the performance indicators will be relevant in all situations, and that several will cut across more than one benchmark (criterion) statement. Furthermore

> "there is not a one-to-one relationship between the benchmarks and the performance indicators since they are pitched at different levels of analysis" (EADTU, 2016, p. 13). (Note: EADTU use the term 'benchmark' for what we refer to as 'criterion').

This observation is true of all the tools we have studied. The fact that many indicators are relevant to more than one criterion, and also that each criterion depends on more than one indicator will affect how these indicators should be used to enable comparisons between courses, programmes or institutions, particularly when stakeholders wish to make comparisons at the level of criteria or higher (e.g. based on several criteria).

The tool that does not use indicators based on ordinal scales is U.S News' "Best online program" tool. This tool has a category entitled "Support services" within its "Student services and technology" category. Criteria are not explicitly described, however a summary of the "Student services and technology" category shows the factors that are considered important by this tool with respect to support services

> "Outside of classes, strong support structures provide learning assistance, career guidance and financial aid resources commensurate with quality campus-based programs" (Eric Brooks & Robert Morse, 2017).

The score for "Support services" is based on the following indicators:

> "student access to 10 equally weighted services: academic advising, bookstore, 24/7 tech support, financial aid services, live librarian, local area network, mentoring, live tutoring, writing workshops, career placement assistance" (*ibid*.).

The information is collected via a questionnaire sent to institutions which run online programs. It appears that the indicators listed above are binary categorical indicators e.g. 'yes'/'no' dependent on ether the institution in question provides each service.

4.1.2 Product (Ossiannilsson *et al.*'s meaning)

In section 3.3 we mentioned Gibbs (2010) discussion of three groups of variables related to the quality of online education, i.e. process variables (those that characterise what is going on in teaching and learning), product variables (those that characterise the outcomes of the educational processes) and presage variables (all the others). Gibbs remarks that presage and product variables cannot explain the variation between institutions in relation to educational gains (though they can explain variations in performance). Measures of educational product such as grades, retention and employability do reflect presage variables

such as funding and reputation, but largely because the best students compete to enter the best-funded and most prestigious institutions and the quality of students is a good predictor of products. The best predictor of educational gain is measures of educational process, i.e.

> what institutions do with their resources to make the most of whatever students they have (Gibbs, 2010, p. 5).

The online version of Matrix 3 shows the variety of indicators that are associated with Gibbs' descriptors. It is our impression that the tools that can be used in a formative way by institutions intending to improve their online provision have a greater focus on indicators related to measurements of the educational processes involved. These tools use ordinal scales as measures of the characteristics of the educational process for indicators within e.g. the 'Course delivery', 'Course design' and 'Curriculum design' categories. For example the E-xcellence tool has a 'Course delivery' criterion which states

> "E-learning systems provide a choice of online tools which are appropriate for the educational models adopted and for the requirements of students and educators".

As mentioned in our discussion of the 'Services' category, in the E-xcellence tool criteria are measured on a five point ordinal scale from 'Excellent in all aspects' to 'Not adequate in majority of aspects'. If, as Gibbs states, measures of educational process are the best predictor of educational gain, then reliable ways of producing indicators of 'process' will be necessary. If indicators based on those within the current set of existing tools are to be used, then a method for enable reliable and repeatable assignment of these ordinal measures will be required.

Some tools include numerical data captured via questionnaires filled in by institutions. Care has to be taken when interpreting this data. For example, the number of students in a class may be important, however, the context in which a 'class' exists will have an impact on what that figure means in terms of is effect on the educational process:

> "In higher education what may matter most is not the size of the largest lecture that is attended on any particular course but the size of the smallest seminar group or problem class that they attend within the same course" (Gibbs, 2010, p. 20).

4.1.3 Management

This category includes criteria and indicators related to considerations such as Institutional strategy, visions, planning and resourcing. Include evaluation and assessment of services and products. Simple indicators such as the presence of relevant policies and plans may be useful, but judging the quality of policy or plans may be more difficult to do.

4.1.4 'Other'

Matrix 1 shows that three of the tools we studied include indicators which do not appear to match the categories identified in Ossiannilsson *et al.*'s study. The set that does not match the categories identified by Ossiannilsson *et al.*'s includes 15 of U-Multirank's teaching and learning indicators. The U-Multirank indicators that do not match includes the following indicators, and for each we have indicated which of Biggs 3Ps they are associated with:

- Graduation rate (for each degree level) [Product]

- Graduating on time (for each degree level) [Product]
- Unemployment rates of graduates [Product]
- Percentage of academic staff with doctorates [Presage]
- Levels of study available [Presage]
- Degree level focus (Number of master and doctorate degrees awarded as a percentage of total number of degrees awarded) [Product]
- Scope (The number of broad educational subject fields in which students have graduated in the latest year available). [Product]

None of these indicators are process indicators, and most are 'product' indicators (in the Biggs/Gibbs use of the term), i.e. they relate to the outcomes of educational processes.

We identified two other tools that include 'other' indicators amongst the online education quality assurance tools we studied. The U.S. News tool features a criterion called 'Peer reputation'. This measure relates to findings from a survey of representatives of 'peer' institutions, who give their opinion of the institution being evaluated, via a questionnaire. Deans and top distance learning officials of schools with online bachelor's programmes are invited to rate other online bachelor's degree programs listed on the survey on a scale of 1 (marginal) to 5 (outstanding) or by responding "don't know" for any programme with which they were unfamiliar (Eric Brooks & Robert Morse, 2017). It is conceivable that participants are asked to make their ratings with direct reference to one or more of Ossiannilsson *et al.'s* categories of indicators but this is not clear in the explanation available. For this reason we have classified this peer rating as "other".

The AQU tool has two categories of criteria which we feel do not wholly match the categories identified in Ossiannilsson *et al.*'s study. The first relates to 'internal assessment' of the process carried out by a programme in preparation for external validation of its performance. In terms of the 3P classification we have classified all the relevant indicators as 'presage' indicators. The second category relates to results relating to both academic and professional achievements, i.e. these are 'Product' indicators. Of the 14 indicators in this category, 9 can be related to Ossiannilsson *et al.*'s categories in particular 'Course delivery', but the other 5 do not.

4.2 Indicator production and use

A variety of processes are used to collect data which are then used to generate the indicators. These include

- Analysis of publicly available data (e.g. publications, staff numbers)
- Self-assessment of aspects institutional and/or course and/or staff performance by the members of staff from the institution, usually guided by a structured rubric or questionnaire. A scoring mechanism may be used to generate an overall grade for the institution or course performance.
- Organisations offering certification or producing rankings collect statistical and other data from institutions, experts and students. This data is then processed to produce one or more indicators, and then to yield the ranking or certificate.
- Assessment of aspects of institutional and/or course and/or staff performance by nominated external reviewers This can take a variety of forms e.g.
    - a visit to the institutions premises

- completion of a questionnaire or other form of structured document.

The indicators in a given scheme may be generated by one or more of the above processes. Matrix 4 summarises the data collection and indicator generation processes used by each of the online education quality assurance tools we studied, plus U-Multirank.

Organisation	Tool	Scope	N	Data sources and processes	Uses
AQU/UOC	Evaluation of online teaching & learning	P + C	94	Self-assessment via scorecards and commentary supported by guidance documents. Each indicator is scored on a 4 point ordinal scale e.g. Not at all positive, Not very positive, Positive, Very positive. This self-assessment is followed by a process of expert review for endorsing and certifying the scores that an institution awards itself.	Certification, Benchmarking, Advisory
EADTU	E-xcellence	P + C	341	Self-assessment via an online questionnaire, in which indicators are to be scored on a five point ordinal scale: Not adequate in some aspects, Not adequate in majority of aspects, Adequate, Adequate with some examples of excellent performance, Excellent in all aspects. This self-assessment can be the basis for a second phase i.e. a “full assessment”) in which expert reviewers visit the university and do an on-site assessment. To receive the E-xcellence Associations Label the institution must integrate the relevant benchmarks into its internal QA-system, thus guaranteeing continuous and repeated use of the E-xcellence benchmarks.	Certification, Benchmarking, Advisory
EADTU	OpenupEd Quality Label	I + C	32	The process (and the indicators) are based on the E-xcellence framework, but using a modified set of indicators appropriate for MOOCs. The overall process (self-assessment followed b expert review is the same as for E-xcellence).	Certification, Benchmarking, Advisory
EFQUEL	ECBCheck	P + C	50	This is a self-assessment system, but the product of the self-assessment may be peer-reviewed to enable the institution to gain a certificate. Each indicator is scored on a 4 point ordinal scale: not met, partly met, met adequately, met excellently	Certification
OLC Online Learning Consortium	Administration of online programs	P + C	75	Self-assessment via scorecards, supported by guidance documents. Each indicator is scored on a 4 point ordinal scale i.e. Deficient, Developing, Accomplished, Exemplary. This self-assessment can be followed by a process of peer review for endorsing and certifying the scores that an institution awards itself.	Certification, Benchmarking, Advisory
Quality Matters	Quality Matters	C	43	Self-assessment via an annotated rubric which guides the user towards what to assess to score each indicator. Each indicator may be scored with either 1, 2 or 3 points although this is implicitly an ordinal scale. The total number of points available from all the 43 indicators is 99. A score of 85% (with Essential Standards being met) qualifies a course to receive a QM Certification for quality course design.	Certification, Benchmarking, Advisory
U.S. News	Best online program	I + P	15	Two sources of data are used. (1) Statistical data collected from institutions via a survey (2) a separate peer reputation survey administered for U.S. News by a market research firm. Academics were ask to rate the academic quality of the other online bachelor's degree programs listed on the survey on a scale of 1 (marginal) to 5 (outstanding) or by responding ""don't know"" about any program with which they were unfamiliar.	Benchmarking
U-Multirank	U-Multirank	I + P	34	(1) Statistical data collected from institutions and programmes via surveys (2) Survey of students (3) Bibliometric and patent data. In the UK and US some of the data requested from institutions can be acquired from national datasets (e.g. those manged by HESA in the UK).	Benchmarking

Matrix 4 Summary of the data collection and indicator generation processes

5 A2.3. Anchoring and weighing current quality criteria to different kinds of online education quality assurance tools

As we mentioned in section 2.1 of this document, we use the term 'criterion' to signify a characteristic of online provision that might be quality assured, the term 'category' to signify a group of criteria, and the term 'indicator' to signify a measure that contributes to, or forms, a measurement of a particular criteria.

In this section we present a methodology for comparing QA tools and their associated indicators for online education. We do this by suggesting ways of appraising criteria and the associated indicators for their relevance, reliability and validity, and by suggesting processes for prioritising (weighting) indicators from the perspectives of different stakeholders.

5.1 Methodology for comparing and evaluating quality criteria and QA tools for online education

Our methodology has considerations for aspects of indicator use, indicator content, and indicator production. It consists of answering a series of questions for each indicator related the following characteristics.

1. Characteristics of indicator content.
   These characteristics relate to what is evaluated and measured by indicators, i.e. the indicators themselves and the associated categories criteria.
2. Characteristics of indicator production.
   These characteristics relate to how data is gathered, and the effect of different data types.
3. Characteristics of indicator use.
   These characteristics relate to how indicators are made use of by a variety of stakeholders.

We begin with an appraisal of indicator content because the intended content of an indicator will determine both how it can be used and how it should be produced. We note however that some stakeholders may well have their own individual perspectives on the relative priority and importance of consideration of these characteristics. For example, media organisations and university management may be more concerned about the workload involved in producing a set of indicators in comparison with their potential to produce newsworthy headlines as well as or instead of their usefulness to prospective students.

5.1.1 Questions related to Indicator Content

In the preceding sections of this deliverable we have broken our analysis down into the three broad categories of services, products and management identified by Ossiannilsson *et al.* in their analysis of the state of the state of the art in quality models in online education (Ossiannilsson *et al.*, 2015). Gibbs (2010) made a distinction between measures of educational performance (e.g. grades achieved) and measures of educational gain (increase in achievement compared to achievement level on entry). To generate an overall view of the quality of the educational experience provided by a particular educational context both educational gain and performance need to be considered, but the relative importance of each of these will vary for different stakeholder groups and for individuals within those groups. However,

> "because educational performance is predicted by the entry standards of students, to compare institutional performance in a valid way it is necessary to measure educational gain: the difference between performance on a particular measure before and after the student's experience of higher education" (Gibbs, 2010, p. 6)

The best predictor of educational gain is measures of educational process, i.e.

> what institutions do with their resources to make the most of whatever students they have (Gibbs, 2010, p. 5).

We note that it is not straight forward to measure educational gain, and discuss the difficulties in section 6.

For any potential indicator, we propose that the questions that should be asked are as follows.

- Does the category represent an aspect of the online teaching and learning process that has been shown by research evidence to influence educational gain?
- If so, what does research evidence say about the amount of influence the category has?
  If research specifies effect sizes related to particular indicators within categories, the effect sizes should be used to inform the relative weighting of the category and its associated indicators.
- Can the type of data used to represent the indicator be used to make valid comparisons between institutions, programmes or courses?
  E.g. if the value of an indicator is represented by nominal or ordinal or data, it may be necessary to include reliability checks to ensure that indicators produced by different individuals or organisations are truly comparable.

5.1.2 Questions related to Indicator use

The IREG Observatory on Academic Ranking and Excellence has published a set of guidelines for Stakeholders of Academic Rankings (IREG Observatory, 2015). These guidelines are intended to provide recommendations for appropriate interpretations, uses and applications of rankings by potential stakeholders, including students and parents, institutions of higher education, policymakers, quality assurance and funding organizations, employers and the media. These guidelines include some general points, i.e. that users should: (a) be clear what academic rankings measure (b) Use academic rankings as just one source of information (c) Pay less attention to precise positions and annual changes: take a long-term view of rankings (d) Carefully read and understand all methodologies. (ibid). Points (a) and (d) relate to the transparency, usability and comprehensibility of the systems, and are applicable to our investigation.

- Is the system which utilises the categories, criteria and indicators usable by the stakeholders it is targeted at?
  A quality system may be complex to use so even if it provides valid comparisons, these may not be usable by all stakeholders. We can use sub-questions related to this theme i.e.
    - Are the system's representation of categories, criteria, indicators and their comparisons comprehensible to the stakeholders it is intended to serve?

    - Are the system's representation of categories, criteria, indicators and their comparisons useful to the stakeholders it is intended to serve?
      For example, is it useful to media organisations, policy makers, and prospective students?
    - Are different types of representations and comparisons needed for each stakeholder group?
- Is the system of categories, indicators and criterion easily modified in the light of new research on the effectiveness of online education which may change the view on particular indicators and weightings?
- Is the set of categories, indicators and criterion easily modified to take account new developments in online education?

5.1.3 Questions related to Indicator production

We have already discussed a range of factors affecting indicator production in section 3.2.3. In addition to the questions we raised in that section we add the following.

- Is the data required to produce all the indicators relevant to a particular category or criterion available from the range of institutions, programmes or courses that are of interest to stakeholders? (If it is not, then the range of comparisons that can be achieved will be limited),
  And is it available from the same scope of analysis? (E.g. institution, programme, course.)
- Is the workload to produce the indicator acceptable to the stakeholder who will do the work?
- Are personnel with the skills required to generate and process data from the sources, and of the types required?
  Different skillsets are required to analyse quantitative data compared to assessing and discussing quality reports and questionnaires.
- Is the rate of change of the category being assessed greater than the frequency of updating of published indicators?
  If the category (characteristic being measured) changes more often than indicator data is collected and processed it will mean that any comparisons based on the indicators are likely to become out of date and lessen their potential usefulness to stakeholders. Many rankings are published annually, and whilst this is a reasonable frequency for many characteristics of online learning, discrepancies may occur. For example, changes in institutional, course or programme performance may occur after that year's indicator data that has already been acquired. These changes may be due to factors such as staff turnover, adoption of new technology platforms, or changes in institutional policies.

6 Discussion and conclusions

In this section we present a summary of important issues, and outline how this deliverable will contribute to future work in CODUR including deliverable IO1.A3 (led by ITD-CNR). Within the following discussion we often refer to U-Multirank as it is the tool which is targeted to adopt the outcomes of the CODUR project. However many of our findings are relevant to other ranking and quality assurance tools, and we have endeavoured to make it clear those that relate to U-Multirank alone.

In this deliverable we have analysed a range of quality assurance and ranking tools for online education, along with a variety of relevant literature, leading to the development of a methodology for systemic comparisons of current online education quality assurance tools and systems. We have referred to evidence which shows that measures of educational processes are better indicators of educational quality than measures of educational performance because "educational performance is predicted by the entry standards of students" (Gibbs, 2010, p. 6). We therefore focused our attention at indicators of the quality of educational processes.

We found that in existing tools, indicators of the quality of educational processes are typically measured using ordinal scales such as "'Deficient', 'Developing', 'Accomplished', 'Exemplary'" or "'Excellent in all aspects', 'Adequate with some examples of excellent performance', 'Adequate', 'Not adequate in some aspects', 'Not adequate in majority of aspects'. We observe that while these scales can be useful to record how a course's (or institution's or programme's) processes may relate to a perceived ideal, and hence be useful for formative development of the course (or institution or programme), it may be difficult to use values of these scales to generate valid and reliable comparisons of courses (or programmes or institutions). This is because human judgement is needed to assign these values, and perceptions of what is "Deficient" or "Accomplished" may vary between judges and also across educational contexts. In its future work, the CODUR project needs to evaluate if comparisons based on ordinal measures are acceptable to the various stakeholders involved. For example, in terms of the workload required to achieve inter-rater reliability to U-Multirank (or similar ranking organisations) and institutions supplying data, and to prospective students in terms of their perceptions of the value of indicators and comparisons based on ordinal scales. Workload has been reported to be an issue by some institutions, e.g.

> "responding to the U.S. News & World Report rankings questionnaire can be a cumbersome process involving upwards of 50 hours of combined effort across various departments" (Reinventing Higher Education, 2016).

Comparison based on product (in Biggs/Gibbs usage) indicators such as graduation rates may be useful to some stakeholders, e.g. media organisations or institutions' marketing departments. However, some care needs to be taken in both using and generating these types of indicators for institutions whose primary delivery method is online education. This is because the characteristics of their student population usually means that there are more variations in terms of study patterns than for the student population of face-to-face universities. However product indicators are potentially easier to generate than process indicators, because often they can be calculated from existing data.

There is ongoing research exploring a variety of methods of assessing learning gains (e.g. 'distance travelled' in learning by students over the duration of their degree). However lack of valid and reliable measures that could be applied systematically across the higher education sector (McGrath, Guerin, Harte, Frearson, & Manville, 2015) means that such

measures are not yet ready to be applied in a tool such as U-Multirank. Simple approaches such as comparing grades on entry vs 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, Zahner, & Benjamin, 2015). However there is work on more sophisticated approaches that may be applicable to ranking and comparison systems in the future, e.g. the Affective-Behaviour-Cognition model (Rogaten, Rienties, Whitelock, Cross, & Littlejohn, 2016).

As Vught and Ziegele of the U-Multirank project have declared, the aim of U-Multirank is to give users choices about how rankings and comparisons are generated:

> "Based on the epistemological position that any choice of sets of indicators is driven by their makers' conceptual frameworks, we suggest a user-driven approach to rankings. Users and stakeholders themselves should be enabled to decide which indicators they want to select to create the rankings that are relevant to their purposes" (Vught & Ziegele, 2011).

However, the approach taken by U-Multirank is not without its critics. In an analysis of ranking systems, Lynch (2015) says of U-Multirank:

> "The focus is on getting the rankings correct, even though the task of ranking incomparable institutions on multiple criteria across different countries and continents is ethically questionable, empirically challenging, and, arguably of primary value to the wealthier students" (Lynch, 2015, p. 201).

In his analysis of IREG's "Berlin Principles on ranking of higher education institutions" (IREG Observatory, 2006) Barron discusses the fifteenth principle

> "Provide consumers with a clear understanding of all of the factors used to develop a ranking, and offer them a choice in how rankings are displayed"

and observes

> "Engaging with U-Multirank becomes a learning experience, rather than a passing judgment. This defeats the efficiency and clarity of rankings in that the user not only has to think heavily on matters of complexity and importance, but also to practice using the tool and to learn how to interpret the results. U-Multirank is exemplary of addressing the fifteenth principle, but is costly, time intensive, and defeats the efficiency that gives rankings their purpose" (Barron, 2017).

Whilst dealing with Lynch's criticisms is beyond the scope of the CODUR project, it is essential that the usability issues raised by Barron are not ignored as work progresses towards establishing the final version of the new online dimension with its adapted indicators. There is a need for more and better information to allow potential students to make informed choices

> "While the privatization of the cost of higher education has advanced enormously, the student/consumer has not been empowered to a similar extent to make smarter choices. The availability and quality of information have simply not improved sufficiently to allow students to make smart choices" (Van Damme, 2015).

One way of increasing usability and stakeholder choice would be for the U-Multirank system to make its data available publicly in an open and machine readable format such as Linked Open Data. This would enable any interested parties to construct their own sets of indicators and representations from the data. Furthermore, it is possible that institutional Linked Open Data repositories could act as sources from which data necessary to generate indicators is harvested by a system such as U-Multirank. We recognise that there is significant effort involved in either of these approaches. However, if either or both were to be mandated by relevant bodies such as the E.U then it would enable the possibility of custom indicator generation queries being specified and run across data from all European universities.

Many of the ranking schemes for face-to-face universities have a focus on research. However. there have been studies which show that excellence in research is not related to excellence in teaching e.g. Marsh and Hattie's meta-analysis "strong support for the typical finding that the teaching-research relation is close to zero" (Marsh & Hattie, 2002) (Considerations of this kind have led to the introduction of the Teaching Excellence Framework in the UK; this runs independently of the Research Excellence Framework). March and Hattie's findings relate to research in general (i.e. research about art, physics or any domain). Research related to teaching, and online teaching in particular, may well improve education quality in institutions where such research occurs. In this case, the effectiveness of the process of adoption of research findings will be an important factor affecting educational quality.

Finally, we reflect on Amsler and Bolsmann's contention

> "The assertion that 'rankings are here to stay' is not an objective representation of reality. It is a politicised speech act rather than a truth claim; a claim on reality, and often a means of precluding critique" (Amsler & Bolsmann, 2012, p. 291).

In practical terms, the benefits gained from a ranking system must be sufficient to merit the investment made by stakeholders in maintaining the system, and for the workload in producing and analysing the data. We observe that the UNIQ scheme reported on by UOC (2015) is no longer in operation

> "After almost ten years, troubles in getting a sustainable model for future development have resulted in the termination of its activities. However, the certification protocols are still valid" (UOC, 2015).

We conclude that reliable and valid indicators are not necessarily sufficient for a sustainable system. Workload, usability, and in general perceived benefit to all stakeholders will all play a part in determining the sustainability of any ranking or comparison system for online universities.

7 References

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

Barron, G. R. S. (2017). The Berlin Principles on Ranking Higher Education Institutions: limitations, legitimacy, and value conflict. *Higher Education, 73*(2), 317-333. doi: 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.

Brooks, E., & Morse, R. (2017). Methodology: Best Online Bachelor's Programs Rankings. Retrieved 24/5/2017, from https://www.usnews.com/education/online-education/articles/bachelors-methodology

Brooks, E., & Morse, R. (2017). Methodology: Best Online Bachelor's Programs Rankings. Retrieved 2/6/2017, from https://www.usnews.com/education/online-education/articles/bachelors-methodology

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.

EADTU. (2016). *Quality Assessment for E-learning: a Benchmarking Approach* K. Kear & J. Rosewell (Eds.), Retrieved from http://e-xcellencelabel.eadtu.eu/images/E-xcellence_manual_2016_third_edition.pdf

Gibbs, G. (2010). *Dimensions of quality*. York: Higher Education Academy

Hood, N., & Littlejohn, A. (2016). MOOC Quality: the need for new measures. *Journal of Learning for Development-JL4D, 3*(3). http://www.jl4d.org/index.php/ejl4d/article/view/165

IREG Observatory. (2006). Berlin principles on ranking of higher education institutions. Retrieved 23/5/2017, from http://ireg-observatory.org/en/index.php/berlin-principles

IREG Observatory. (2015). IREG Guidelines for Stakeholders of Academic Rankings Retrieved 23/5/2017, from http://ireg-observatory.org/en/pdfy/ieg-guidlines-for-stakeholders-of-academic-ranking.pdf

Kear, K. (2011). *Online and social networking communities: A best practice guide for educators*: Routledge.

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

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

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. Luxembourg: Publications Offi ce of the European Union.

McGrath, C. H., Guerin, B., Harte, E., Frearson, M., & Manville, C. (2015). Learning gain in higher education. Cambridge: RAND Corporation/HEFCE. Retrieved from http://www.hefce.ac.uk/pubs/rereports/year/2015/learninggain/.

Moore, M., Lockee, B., & Burton, J. (2002). Measuring success: evaluation strategies for distance education. *EDUCAUSE Quarterly, 25*(1), 20-26.

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. Oslo: International Council for Open and Distance Education (ICDE).

Rogaten, J., Rienties, B., Whitelock, D., Cross, S., & Littlejohn, A. (2016). A multi-level longitudinal analysis of 80,000 online learners: Affective-Behaviour-Cognition models of learning gains.

Sandmaung, M., & Ba Khang, D. (2013). Quality expectations in Thai higher education institutions: Multiple stakeholder perspectives. *Quality Assurance in Education, 21*(3), 260-281.

ShanghaiRanking Consultancy. (2017). ARWU (Academic Ranking of World Universities). Retrieved 18/05/2017, from http://www.shanghairanking.com

Shelton, K., Saltsman, G., Holstrom, L., & Pedersen, K. (2014). *Quality Scorecard Handbook: Criteria for Excellence in the Administration of Online Programs* (pp. 142). Retrieved from https://onlinelearningconsortium.org/consult/quality-scorecard/

Times Higher Education. (2017). World University Rankings. Retrieved 18/05/2017, from https://www.timeshighereducation.com/world-university-rankings

Tsekleves, E., Cosmas, J., & Aggoun, A. (2014). Benefits, barriers and guideline recommendations for the implementation of serious games in education for stakeholders and policymakers. *British Journal of Educational Technology*.

U-Multirank project. (2017). U-Multirank | Universities compared. Your way. Retrieved 22/5/2017, from http://www.umultirank.org/

UOC. (2015). *Making Online Education more Visible and Reputable*. NORA Project Interim Report. Universitat Oberta de Catalunya.

Van Damme, D. (2015). Global higher education in need of more and better learning metrics. Why OECD's AHELO project might help to fill the gap. *European Journal of Higher Education, 5*(4), 425-436. doi: 10.1080/21568235.2015.1087870

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*.

Warburton, S., & Mor, Y. (2015). A set of patterns for the structured design of MOOCs. *Open Learning: The Journal of Open, Distance and e-Learning*, 1-15. doi: 10.1080/02680513.2015.1100070

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. doi: 10.1080/03075079.2015.1004239

Appendix 1. U-Multirank Teaching and Learning indicators

Indicator	Explanation	Ossiannilsson *et al.*'s Categories of indicators
Bachelors graduation rate	The percentage of new entrants that successfully completed their bachelor programme.	
Masters graduation rate	The percentage of new entrants that successfully completed their master programme.	
Graduating on time (bachelors)	The percentage of graduates that graduated within the time expected (normative time)	
Graduating on time (masters)	The percentage of graduates that graduated within the time expected (normative time)	
Graduation rate long first degree	The percentage of new entrants that successfully completed their long first degree programme.	
Graduating on time (long first degree)	The percentage of graduates that graduated within the time expected (normative time) for their long first degree programme.	
Relative BA graduate unemployment	The percentage of bachelor graduates unemployment 18 months after graduation.	
Relative MA graduate unemployment	The percentage of master graduates unemployment 18 months after graduation.	
Relative graduate unemployment long first degree	The percentage of long first degree programme graduates unemployment 18 months after graduation.	
Student - staff ratio	The number of students (headcount) per member of the academic staff (fte). Staff solely involved in research is excluded.	**Course delivery**
Graduating on time (bachelors)	The percentage of graduates that graduated within the time expected (normative time) for their bachelor programme.	
Graduating on time (masters)	The percentage of graduates that graduated within the time expected	

	(normative time) for their masters programme.	
Academic staff with doctorates	The percentage of academic staff holding a doctorate (PhD or equivalent).	
Contacts with work environment (bachelors)	A composite measure representing at bachelor level: (1) the inclusion of internships / phases of practical experience or external projects in the curriculum; (2) the percentage of students doing an internship; (3) teaching by practitioners from outside the university departments; and, (4) the percentage of degree theses made in cooperation with industry/external organisations.	**Course design , Course delivery**
Contact with work environment (masters)	A composite measure representing at masters level: (1) the inclusion of internships / phases of practical experience or external projects in the curriculum; (2) the percentage of students doing an internship; (3) teaching by practitioners from outside the university departments; and, (4) the percentage of degree theses made in cooperation with industry/external organisations.	**Course design , Course delivery**
Hospital beds available for teaching	The number of beds available for teaching in university hospital and affiliated hospitals per 100 students.	**Strategy and management**
Innovative forms of teaching and assessment	The percentage of examinations (in medical doctor training programmes) which use innovative forms of assessment (assessment of practical work by faculty and structured clinical cases).	**Curriculum design , Course design , Course delivery**
Overall learning experience	An assessment of the quality of the overall learning experience, based on a survey of the students.	**Curriculum design , Course design , Course delivery**
Quality of courses & teaching	An assessment of the quality of teaching provision, based on a student satisfaction survey.	**Curriculum design , Course design , Course delivery**
Organisation of program	An assessment of the organisation of the programme, based on a student satisfaction survey.	**Curriculum design**

Contact with teachers	An assessment of the feedback given by teachers, based on a student satisfaction survey.	**Course delivery**
Inclusion of work/practical experience	An assessment of the inclusion of work experience and of elements related to work practice, based on a student satisfaction survey.	**Course design , Course delivery**
Library facilities	An assessment of the quality of library services for students, based on a student satisfaction survey.	**Course delivery**
Laboratory facilities	An assessment of the quality of laboratories available to students, based on a student satisfaction survey.	**Course delivery**
IT provision	Student assessment of the quality of IT services for students, based on a student satisfaction survey.	**Course delivery**
Room facilities	An assessment of lecture halls and seminar rooms, based on a student satisfaction survey.	**Course delivery**
Inclusion of practical experience/clerkships (medicine)	The integration of practical experience with patient contact into the study programme, based on a student satisfaction survey.	**Curriculum design , Course design , Course delivery**
Bedside teaching	An assessment of bedside teaching concerning mentoring, suitability of rooms and variety of diagnostic techniques applied, based on a student satisfaction survey.	**Curriculum design , Course design , Course delivery**
Linking clinical/preclinical teaching	The integration of pre-clinical/theoretical and clinical courses, based on a student satisfaction survey.	**Curriculum design , Course design , Course delivery**
Skills Labs	An assessment of the skills labs and training centers concerning maintenance, accessibility, technical facilities and mentoring, based on a student satisfaction	**Course delivery**
Expenditure on teaching (%)	Percentage of total institutional expenditure dedicated to teaching activities	**Strategy and management**
Degree level focus	Number of master and doctorate degrees awarded as a percentage of total number of degrees awarded.	

Scope	The number of broad educational subject fields (ISCED97) in which students have graduated in the latest year available.	
Level of study	The degree levels at which the institution awards degrees	