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

codur

Creating an Online Dimension for University Rankings

PROJECT DELIVERABLE: IO2.A1

Design, test, valorize and refine the toolbox for data gathering and testing the new ranking tool with European and worldwide online university partners

Francesca Pozzi (CNR-ITD), Flavio Manganello (CNR-ITD), Marcello Passarelli (CNR-ITD), Donatella Persico (CNR-ITD), Andrew Brasher (OU), Denise Whitelock (OU), Wayne Holmes (OU), Maria Taulats (UOC), Lídia Toda (UOC, Albert Sangrà (UOC)

Date: 29/03/2018

Version 1.3

Universitat Oberta de Catalunya
The Open University
Consiglio Nazionale delle Ricerche

Table of contents

List of Figures

List of Tables

1 Introduction

The CODUR project acknowledges the gap between, on the one hand, the wide usage of existing university ranking systems and, on the other hand, the lack of specific criteria and indicators addressing the online component of universities, although online universities are known to play a crucial role in European Higher Education (Brasher, Holmes, & Whitelock, 2017; European Commission, 2014). Consequently, the ultimate goal of the CODUR project is to propose a set of criteria and indicators specifically devoted to the evaluation of online institutions that should be then integrated with already existing rakings systems. Specifically, the CODUR proposal refers to the EU supported U-Multirank system (Van Vught & Ziegele, 2011), because its multidimentional approach is particularly fitting to the diversified needs of online learners.

To achieve this goal, during the first year the main tasks of the project had to do with the design and definition of a set of criteria and indicators. In particular, the project envisaged three main tasks:

- IO1-A1: Aligning the provision of online Higher Education Institutions worldwide with U-Multirank categories.
- IO1-A2: Designing the means for systematic comparisons of current online education quality assurance tools and systems.
- IO1-A3: Develop, test and refine representative performance online quality education indicators based on common criteria.

During the second year, instead, the project main tasks have to do with the production of guidelines for integrating online education quality assurance metrics in existing ranking systems, such as the EU supported U-Multirank. More in detail, according to the proposal, the project envisages the following three main tasks:

- IO2-A1: Design, test, valorize and refine the toolbox for data gathering and testing the new ranking tool with European and worldwide online university partners
- IO2-A2: Create a set of new guideline for adapting CODUR indicators and toolbox resources, under the supervision and feedback of a U-Multirank expert
- IO2-A3: Guidelines to improve the performance of online institutions.

In this deliverable, we describe the main activities carried out under IO2-A1 and their main outputs. In particular, the document is structured as follows: in Section 2, we describe the Second Round of a Delphi Study aimed to define the CODUR criteria and indicators. The First Delphi Round, carried out during the first year of the project and reported in IO1-A3, contributed to the production of a preliminary list of criteria. The Second Delphi Round is based on such list and aims to identify the final list of criteria and indicators, which will constitute one of the main CODUR output (described in Section 3).

In Section 4, we describe the way we have developed a toolbox, aimed to data gathering and testing of the new set of criteria and indicators.

In Section 5, we describe how we have used such toolbox with some supporting online Universities, which have been engaged in the test phase.

In Section 6, we discuss the main lessons learnt and provide recommendations for future work.

2 Refining the CODUR criteria and indicators

In CODUR, the process of choosing and refining criteria and indicators for quality assessment of online Higher Education Institutions is based on consultation with experts in the field and entails a Delphi Study and a face-to-face workshop involving selected informants. As already mentioned, the first round of the Delphi Study and the workshop, both carried out during Year I, led to identify the set of criteria to be considered, their sub-criteria, and their relative weights[1]. The methods and results of the first round of the Delphi Study, as well as of the workshop, are reported in the deliverable of Task IO1-A3 (Pozzi, Manganello, Passarelli, & Persico, 2017).

To produce a solid base for the second round, we took on board the comments and feedback coming from the previous activities and produced a revised set of criteria. Sub-criteria were subsequently operationalized, so as to propose a list of observable indicators to be considered for inclusion. Thus, the main objective of the second round of the Delphi Study was to evaluate the observable indicators and to inform the selection of a limited number of them, with the eventual aim of obtaining a short list of measurable indicators that would balance validity with ease of measurement[2]. A second objective was to obtain estimates of weights for these observable indicators, so that the final ranking system can consider both the relative importance of criteria, and the relative importance of the internal aspects of each criterion.

The following sections detail the implementation and results of the second round of the Delphi Study.

2.1 Method: the Delphi Study

As already mentioned in the deliverable of Task IO1-A3, the Delphi method is a research technique based on consultation with a panel of experts, first described by Dalkey & Helmer (1963). Delphi studies are typically used as forecasting tools, with the aim of predicting trends, problems and possible developments of a given sector (Rowe & Wright, 1999). However, Delphi studies can also be used as tools for analysis based on experts' judgment.

The Delphi method, along with the main phases typically envisaged, is described in detail in the deliverable of Task IO1-A3 (Pozzi et al., 2017). However, before moving on to the description of the CODUR Delphi, we summarize below the method's main characteristics.

Delphi is a research technique based on consultation with a panel of experts, which is typically used to forecast future scenarios or to analyze complex issues by drawing on the expertise of key stakeholders. Delphi studies have several variations, but they are based on an iterative process of consultation with a panel of experts, where at each iteration (except the first) the panelists are provided with a synthesis of outcomes from the previous round. The participants have to remain anonymous to each other to limit the risk of data bias deriving from negative dynamics between experts known to each other. The idea is to collect experts' opinions on the same topic on the assumption that participants will revise and refine their position based on the outcome of previous rounds. Indeed, in Delphi, participant responses tend to change over time, either because the experts' opinions converge or because outcomes from previous

[1] The terms criteria, sub-criteria and (observable) indicators, in CODUR, identify respectively broad aspects to be evaluated (e.g. quality of teaching and learning), facets of these aspects (e.g. course quality) and metrics that can be used to measure them (e.g. students satisfaction of learning material as measured through astudent survey).

[2] The need for reducing the number of operationalized indicators to a manageable number was clearly expressed by U-Multirank representatives participating to the second CODUR meeting, just before the second Delphi round started.

rounds bring new issues and points of view to their attention, thus triggering further reflection. The Delphi study ends when the experts' responses become stable across rounds and/or the reserachers are satisfied with the depth of the results obtained.

2.1.1 The CODUR Delphi

Within the CODUR project, the Delphi Study was implemented as a tool for analysis, rather than forecasting. The aim of the CODUR Delphi Study is to remotely involve a worldwide community of experts and informants, with the aim of evaluating the criteria and observable indicators for assessing the quality of online Higher Education Institutions. The Delphi Study structure also allowed participants to suggest indicators that were not initially considered.

The CODUR Delphi Study research design is represented in the diagram below (Figure 1). The methodology for carrying out round 1, as well as its results, are detailed in the project deliverable for Task IO1-A3. The present deliverable will only consider round 1 in relation to how it impacted on the preparation of round 2, i.e. how its results informed the goals and structure of the second round of the Delphi Study.

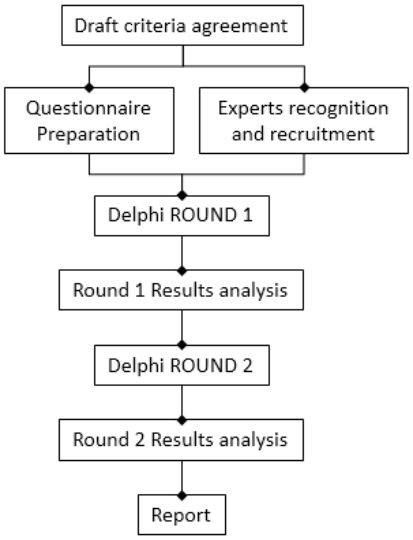

Figure 1 - CODUR Delphi Study Research Design.

In preparation for round 1 of the Delphi Study, nine criteria for the assessment of quality of online Higher Education Institutions were identified by the project partners (quality of teaching, quality of the learning experience, quality of student support, quality of teacher support, reputation/impact, quality of research, quality of organization, sustainability of the institution, quality of the technological infrastructure). This list of criteria was used as a starting point for the iterative process that would form the core of the Delphi study.

2.1.1.1 Round 1 of the CODUR Delphi

The first round of the Delphi Study, carried out through an online survey, focused on: (1) identifying the relevant criteria for assessing the quality of online Higher Education Institutions; (2) finding a non-ambiguous, agreed-upon definition for each criterion; (3) identifying the sub-criteria that form each criterion, a necessary preliminary step to the operationalization of each criterion; (4) estimating relative weights to be applied to the criteria.

Goal (1) was met by open-ended questions in which participants were free to add, remove, merge, or split criteria from the proposed list. No additional criteria were proposed by the participants. However, some participants suggested to merge quality of teaching and quality of

the learning experience in a single criterion. This position was echoed during the EMEM workshop that followed the Delphi, in which the same suggestion was put forward by participating stakeholders (see Task IO1-A3 for more details about round 1). In the round 2 Delphi questionnaire, these two criteria were merged into the criterion "quality of teaching and learning".

Goal (2) was met by presenting participants with our definition for each criterion, and asking them to comment it and suggest amendments. The definitions were generally accepted, and suggestions were usually limited to rewordings that were integrated in the criteria definitions used in round 2.

Goal (3) was met using questions aimed to collect both quantitative and qualitative data. For each criterion, participants were presented with a list of proposed sub-criteria, and they were asked to rate how important they deemed them to be, on a scale from 0 (not at all important) to 4 (extremely important). After rating all the proposed sub-criteria for a criterion, participants were asked to suggest the addition and/or removal of sub-criteria from the list. Results for this goal led to considerable additions to the lists of sub-criteria, which were carefully considered when devising observable indicators for round 2 of the Delphi Study.

Finally, goal (4) was met by asking participants to rank the criteria in order of importance. The rankings were analyzed using Thurstone Scaling Type V, which provided relative weights that were successfully replicated on a separate sample (see Pozzi et al., 2017, Section 4.2).

Results from the first Delphi round were analyzed by the ITD-CNR team, with the objective of creating a list of proposed observable indicators for each of the criteria. Following the participants' suggestion, two criteria (quality of teaching and quality of the learning experience) were merged and, for each of the remaining eight criteria, the importance rating of sub-criteria was considered and led to the removal or merge of some sub-criteria. Lastly, observable indicators were proposed for each sub-criterion. These formed the main object of the second round of the Delphi Study.

2.1.1.2 Round 2 of the CODUR Delphi

The second round of the Delphi carried out during Year 2 of the project had two main goals: guiding the selection of the best indicators for each criterion, so as to obtain a measure that is both valid and efficient, and a relative weight to be applied to indicators within each criterion.

Table 1 reports the eight criteria, the revised list of sub-criteria used for devising observable indicators, and the list of indicators proposed to the experts of round 2. For each of the eight CODUR criteria, participants were presented with a list of proposed observable indicators, each listed under the heading of the sub-criteria they aimed at operationalizing. For each criterion, participants were asked to select no more than half of the proposed observable indicators, choosing among the list the indicators they deemed most important. For example, for the criterion 'quality of teaching and learning' participants were presented with twelve observable indicators grouped under eight sub-criteria headings, and were asked to select no more than six indicators in total (with no restrictions on how the selected indicators were distributed between sub-criteria). In case the number of proposed indicators within a criterion was odd, the number of indicators to be selected was rounded down. The order of presentation of criteria, sub-criteria and indicators within each sub-criterion was randomized for each participant.

Table 1 - Criteria, sub-criteria and proposed observable indicators considered in round 2 of the Delphi Study.

Criterion	**Sub-criterion**	**Observable indicator**
Quality of teaching & learning	Quality of overall learning experience	Student satisfaction of the overall learning experience (through student survey)
	Quality of pedagogy /methodology	Student satisfaction regarding adequacy of the adopted pedagogical approaches to the learning objectives (through student survey)
	Quality of course / learning design	Institutional support for learning design (in terms of tools, formats, etc.) (data provided by the institution)
		Percentage of courses that propose personalized paths to reach the learning objectives (for example offering different materials/activities depending on culture, learning style, background, etc.) (data provided by the institution or review by external panel)
		Percentage of courses that support self-regulated learning (e.g. tools for deadline management, tools for progress tracking, self-evaluation tools) (data provided by the institution or review by external panel)
	Quality of learning materials /activities	Student satisfaction regarding learning materials (through student survey)
		Student satisfaction regarding proposed activities (through student survey)
	Quality of assessment	Percentage of courses/examinations that make use of diverse forms of assessment (quantitative and qualitative approaches, human-based and technology-based tools, etc.) (data provided by the institution or review by external panel)
	Tracking of online interactions	Percentage of courses that provide performance reports to learners & teachers by means of learning analytics (data provided by the institution or review by external panel)
		Student and teacher satisfaction regarding performance reports (through student and teacher survey)
	Standards for regulating teacher-student interactions	Existence of suggested standards for feedback provision (e.g. time threshold, ...) (data provided by the institution)
	Quality control and teacher assessment	Frequency of course/programme evaluation (data provided by the institution)

Quality of student support	Quality of interactions between educators and students	Ratio tutors/students (data provided by the institution)
		Student satisfaction regarding interactions with teachers/tutors (through student survey)
	Technology support	Student satisfaction with technology support (including Helpdesk, FAQ, wizards, support material and initial training) (through student survey)
		Helpdesk average response time to students (data provided by the institution)
	Orientation services to help learners taking decisions about their learning path	Student satisfaction with orientation services (through student survey)
Quality of teacher support	Technology support	Teacher/tutor satisfaction with technology support (including Help desk, FAQ, wizards, support material and initial training) (through teacher survey)
		Helpdesk average response time to teachers/tutors (data provided by the institution)
	Opportunities for teaching staff to be trained	Number of hours of training (or equivalent) made available for teachers/ tutors by the institution per year (data provided by the institution)
		Number of hours of training devoted to teaching staff concerning online learning per year (data provided by the institution)
		Number of hours of training devoted to teaching staff concerning code of ethics per year (data provided by the institution)
		Teacher/tutor satisfaction of training opportunities (through teacher survey)
	Support to teaching staff	Teacher/tutor satisfaction with Community building tools made available by the institution (through teacher survey)
		Teacher/tutor satisfaction with design patterns and OER (Open Education Resources) repositories made available by the institution (through teacher survey)
		Teacher/tutor satisfaction with feedback on their courses derived from students' surveys (through teacher survey)
Reputation/Impact	Social impact & responsibility	Percentage of credits given in service-learning activities, in relation to total number of credits. Service

		Learning involves students in community service activities and applies the experience to personal and academic development. Service-learning takes place outside the Higher Education Institution (HEI) (data provided by the institution)
	Communication strategies & Visibility on academic social networks	SEO (Search Engine Optimization) on institutional Websites (data provided by the institution). SEOs are strategies and activities aimed to improve visibility of a website on Internet search engines.
		Position on Webometrics University Ranking (data provided by the institution). The Webometrics Ranking of World Universities, also known as Ranking Web of Universities, is a ranking system for the world's universities.
		Number of clicks/likes/shares/comments/followers/impressions on academic social networks, such as Academia.edu, ResearchGate etc. (data provided by the institution)
	Job opportunities for graduates	Percentage of post-graduated actively engaged after graduation (data provided by the institution)
		Percentage of former students employed in job sectors matching their degree (data provided by the institution)
	International orientation	A composite measure taking into account the existence of joint/dual degree programmes, the inclusion of study periods abroad, the % of international (degree and exchange) students, the % of international academic staff (data provided by the institution)
	Internship and mobility opportunities	The number of student internships (total / per year) (data provided by the institution)
		The number of student mobility (total / per year) (data provided by the institution)
	Relationship with the territory	Percentage of student internships in the region (data provided by the institution)
		The number of theses (BA and MA) with regional organisations (data provided by the institution)
		The proportion of external research revenues - apart from government or local authority core/recurrent grants – that comes from regional sources (i.e. industry, private organisations, charities) (data provided by the institution)

		The percentage of graduates who found their first job (after graduation) in the region where the university is located (data provided by the institution)
	Representation on national forums	The number of national forums joined (data provided by the institution)
Quality of research	Research in online teaching & learning (research groups, research projects, etc.)	Internal budget devoted to research on online learning and teaching per Full Time Equivalent (FTE) academic staff (data provided by the institution)
		External research income concerning projects on online learning and teaching per Full Time Equivalent (FTE) academic staff (data provided by the institution)
		Number of visiting scholars per Full Time Equivalent (FTE) academic staff (data provided by the institution)
	Teaching staff engaged in research in online teaching & learning	Percentage of Full Time Equivalent (FTE) staff involved in research on online learning and teaching (data provided by the institution)
		Number of doctorate degrees in online teaching and learning per Full Time Equivalent (FTE) (data provided by the institution)
	Research in online teaching & learning - Output	Yearly average n. of publications on online teaching & learning per Full Time Equivalent (FTE) academic staff (WoS or Scopus publications) (data provided by the institution or review by external panel)
	Internationalization	Yearly average number of publications with authors from other countries per Full Time Equivalent (FTE) academic staff (WoS or Scopus publications) (data provided by the institution or review by external panel)
		Percentage of doctorate degrees that are awarded to international doctorate candidates (data provided by the institution)
	Disciplinary research (research groups, research projects, etc.)	Internal budget devoted to disciplinary research per Full Time Equivalent (FTE) academic staff
		External research income concerning disciplinary projects per Full Time Equivalent (FTE) academic staff
		Number of doctorate degrees in disciplinary research
	Disciplinary research - Output	Yearly average n. of publications per Full Time Equivalent (FTE) academic staff (WoS or Scopus publications) (data provided by the institution or review by external panel)

Quality of organization	Credit transfer system aligned with national (and /or European) systems and operates bidirectionally	Credit transfer system adopted by the institution (data provided by the institution)
	Bureaucratic policies able to cater for the needs of e-learning courses	Operations performable online (checklist: subscription, following lectures, examination, vote registration, ...) (data provided by the institution)
	Existence of a complaints and appeals system for learners	Percentage of student complaints or appeals solved/closed (data provided by the institution)
		Average time (days) for processing complaints/appeals (data provided by the institution)
	Bureaucratic support services (providing assistance for admission, financial issues, registration, enrolment, etc.)	Number of full-time equivalents (FTEs) employed for non-instructional, non-technical support services (providing assistance for admission, financial issues, registration, enrolment, etc.) weighted by student satisfaction for the service (data provided by the institution + student survey)
	Structures such as libraries, labs, etc.	Student satisfaction for room, laboratory and library facilities (through student survey)
	Ability of managing time and avoiding workload	Student paperwork / online forms / front office time burden per week (through student survey)
		Academic staff paperwork / online forms / front office / commissions time burden per week (through teacher survey)
	Student satisfaction	Student satisfaction for organization (through student survey)
Sustainability of the Institution	Institutional Strategic Plan (ISP) for online education	Availability of an Institutional Strategic Plan for Online Learning (online vision statement, online mission statement, online learning goals and action steps, ...) (data provided by the institution)
	Overall coherence of program design and provision (interconnections among courses, flexibility of the design, clarity of program design, …)	Percentage of curriculum changes resulting from an assessment of student learning (either formal or informal) within a fiscal year [a measure on increased flexibility within the curriculum development process to better respond to a rapidly changing world] (data provided by the institution)

	Resources (including financial ones) specifically devoted to the online program	Percentage of total institutional expenditure dedicated to online programmes (data provided by the institution)
	Sustainability of the portfolio of programmes	Ratio new students added / lost students per program (in the past year) (data provided by the institution)
		Ratio new students added in the past year / students still in the program (data provided by the institution)
	Clear policy regarding OERs and MOOCs	Percentage of Open Educational Resources used on the total of learning materials (data provided by the institution)
		Percentage of Massive Open Online Courses on the total number of courses offered (data provided by the institution)
Quality of the technological infrastructure	Quality of the overall technological infrastructure	Student satisfaction with the overall learning platform (through student survey)
	Robustness of the learning platform	Composite measure of server error rates, average response times, peak response times, and uptime (through technical institutional survey)
	Flexibility of the learning platform	Checklist of functionalities supported by the platform (through technical institutional survey)
	Adequacy of data security mechanisms	Measure of data security self-assessment (through technical institutional survey).
	Compliance of interfaces and contents with usability and accessibility standards	Measure of compliance with the accessibility guidelines WCAG 2.0 (through technical institutional survey)
	Existence of a plan for system maintenance and contingency management	Availability of a plan for system maintenance and contingency management (data provided by the institution)
	Interoperability of the learning platform	Measure of interoperability (Interoperability with external open sites (e.g., social media, DropBox, Google Drive), interoperability between LMSs (Learning Management Systems), information and teaching/learning materials exchange (LTI, SCORM, ...), Single sign-on (SSO) access control, etc. (data provided by the institution)

Both round 1 and 2 of the Delphi Study were carried out using LimeSurvey, a free and open source online survey application.

2.1.2 Participants

All the experts and informants that took part to round 1 of the Delphi were sent the invitation to participate to round 2. This list of participants includes students, university teachers and researchers – at an international level – who are considered informed and competent on the topic of online higher education.

Twenty-one of the 40 round 1 participants took part to the second round of the Delphi. Of these, 3 identify as researchers, 5 as higher education teachers, and 6 as both educators and researchers. The remaining 7 stakeholders identify as other categories (e.g. retired educators, higher education managers, rector's delegates for e-learning, etc.)

Nine participants reported Italy as their country, while the rest of the sample reported as their country UK (4), Australia (2), or Bulgaria, Canada, Catalonia, The Netherlands, Spain, Turkey (1 each).

2.1.3 Output

In this section, we describe the results of the second round of the Delphi, as those of the first round have already been reported in the deliverable of Task IO1-A3 (Pozzi, Manganello, Passarelli, & Persico, 2017).

Indicators' selection

As mentioned above, participants were asked to select no more than half of the proposed indicators for each criterion. The number of times an indicator was chosen by participants was used to inform the selection of observable indicators to retain. For each criterion, only half of the proposed indicators were retained (rounding down), choosing the ones most often selected by participants. In case of ties at the selection boundary, indicators with the same number of selections were both retained (e.g., if a criterion had 6 proposed indicators, we would have aimed at retaining 3; but if the third and fourth most-selected indicators had the same number of selections, we would retain both of them). This led to the selection of 38 out of 75 observable indicators.

Indicators' weighting

Relative weights for indicators within each criterion were computed as the relative proportion of selections for each indicator on the total number of selections of retained indicators for that criterion. For example, if an indicator had three retained criterions, selected respectively 12, 9, and 6 times, their relative weights would be 12/(12+9+6)=44%, 9/(12+9+6)=33%, and 6/(12+9+6)=22%.

Table 2 reports all the proposed indicators for each criterion, along with the number of times they were selected by participants, whether they were retained or not, and the weight to be applied to them for measuring the criterion.

Table 2 – Number of selections for each proposed indicator and relative weight for measuring their criterion.

Criterion	Observable indicator	Times selected	Retained?	Weight

Quality of teaching & learning	Student satisfaction of the overall learning experience (through student survey)	13	YES	16%
	Student satisfaction regarding adequacy of the adopted pedagogical approaches to the learning objectives (through student survey)	13	YES	16%
	Institutional support for learning design (in terms of tools, formats, etc.) (data provided by the institution)	12	YES	15%
	Percentage of courses that propose personalized paths to reach the learning objectives (for example offering different materials/activities depending on culture, learning style, background, etc.) (data provided by the institution or review by external panel)	9	YES*	11%
	Percentage of courses that support self-regulated learning (e.g. tools for deadline management, tools for progress tracking, self-evaluation tools) (data provided by the institution or review by external panel)	6	NO	-
	Student satisfaction regarding learning materials (through student survey)	12	YES	15%
	Student satisfaction regarding proposed activities (through student survey)	8	NO	-
	Percentage of courses/examinations that make use of diverse forms of assessment (quantitative and qualitative approaches, human-based and technology-based tools, etc.) (data provided by the institution or review by external panel)	11	YES	14%
	Percentage of courses that provide performance reports to learners & teachers by means of learning analytics (data provided by the institution or review by external panel)	6	NO	-
	Student and teacher satisfaction regarding performance reports (through student and teacher survey)	9	YES*	11%
	Existence of suggested standards for feedback provision (e.g. time threshold, ...) (data provided by the institution)	8	NO	-
	Frequency of course/programme evaluation (data provided by the institution)	7	NO	-

Quality of student support	Ratio tutors/students (data provided by the institution)	3	NO	-
	Student satisfaction regarding interactions with teachers/tutors (through student survey)	15	YES	56%
	Student satisfaction with technology support (including Helpdesk, FAQ, wizards, support material and initial training) (through student survey)	12	YES	44%
	Helpdesk average response time to students (data provided by the institution)	7	NO	-
	Student satisfaction with orientation services (through student survey)	5	NO	-
Quality of teacher support	Teacher/tutor satisfaction with technology support (including Help desk, FAQ, wizards, support material and initial training) (through teacher survey)	17	YES	35%
	Helpdesk average response time to teachers/tutors (data provided by the institution)	5	NO	-
	Number of hours of training (or equivalent) made available for teachers/ tutors by the institution per year (data provided by the institution)	8	NO	-
	Number of hours of training devoted to teaching staff concerning online learning per year (data provided by the institution)	11	YES	22%
	Number of hours of training devoted to teaching staff concerning code of ethics per year (data provided by the institution)	0	NO	-
	Teacher/tutor satisfaction of training opportunities (through teacher survey)	12	YES	24%
	Teacher/tutor satisfaction with Community building tools made available by the institution (through teacher survey)	8	NO	-
	Teacher/tutor satisfaction with design patterns and OER (Open Education Resources) repositories made available by the institution (through teacher survey)	4	NO	-
	Teacher/tutor satisfaction with feedback on their courses derived from students' surveys (through teacher survey)	9	YES	18%
Reputation/Impact	Percentage of credits given in service-learning activities, in relation to total number of credits. Service Learning involves students in community	8	YES	11%

	service activities and applies the experience to personal and academic development. Service-learning takes place outside the HEI (data provided by the institution)			
	SEO (Search Engine Optimization) on institutional Websites (data provided by the institution). SEOs are strategies and activities aimed to improve visibility of a website on Internet search engines.	5	NO	-
	Position on Webometrics University Ranking (data provided by the institution). The Webometrics Ranking of World Universities, also known as Ranking Web of Universities, is a ranking system for the world's universities.	5	NO	-
	Number of clicks/likes/shares/comments/followers/impressions on academic social networks, such as Academia.edu, ResearchGate etc. (data provided by the institution)	6	YES*	8%
	Percentage of post-graduated actively engaged after graduation (data provided by the institution)	8	YES	11%
	Percentage of former students employed in job sectors matching their degree (data provided by the institution)	14	YES	19%
	A composite measure taking into account the existence of joint/dual degree programmes, the inclusion of study periods abroad, the % of international (degree and exchange) students, the % of international academic staff (data provided by the institution)	12	YES	16%
	The number of student internships (total / per year) (data provided by the institution)	6	YES*	8%
	The number of student mobility (total / per year) (data provided by the institution)	10	YES	14%
	Percentage of student internships in the region (data provided by the institution)	3	NO	-
	The number of theses (BA and MA) with regional organisations (data provided by the institution)	5	NO	-
	The proportion of external research revenues - apart from government or local authority core/recurrent grants – that comes from regional sources (i.e. industry, private organisations, charities) (data provided by the institution)	9	YES	12%

	The percentage of graduates who found their first job (after graduation) in the region where the university is located (data provided by the institution)	5	NO	-
	The number of national forums joined (data provided by the institution)	1	NO	-
Quality of research	Internal budget devoted to research on online learning and teaching per Full Time Equivalent (FTE) academic staff (data provided by the institution)	13	YES	16%
	External research income concerning projects on online learning and teaching per Full Time Equivalent (FTE) academic staff (data provided by the institution)	7	NO	-
	Number of visiting scholars per Full Time Equivalent (FTE) academic staff (data provided by the institution)	7	NO	-
	Percentage of Full Time Equivalent (FTE) staff involved in research on online learning and teaching (data provided by the institution)	14	YES	18%
	Number of doctorate degrees in online teaching and learning per Full Time Equivalent (FTE) (data provided by the institution)	3	NO	-
	Yearly average n. of publications on online teaching & learning per Full Time Equivalent (FTE) academic staff (WoS or Scopus publications) (data provided by the institution or review by external panel)	14	YES	18%
	Yearly average number of publications with authors from other countries per Full Time Equivalent (FTE) academic staff (WoS or Scopus publications) (data provided by the institution or review by external panel)	8	YES*	10%
	Percentage of doctorate degrees that are awarded to international doctorate candidates (data provided by the institution)	2	NO	-
	Internal budget devoted to disciplinary research per Full Time Equivalent (FTE) academic staff	8	YES*	10%
	External research income concerning disciplinary projects per Full Time Equivalent (FTE) academic staff	9	YES	11%

	Number of doctorate degrees in disciplinary research	4	NO	-
	Yearly average n. of publications per Full Time Equivalent (FTE) academic staff (WoS or Scopus publications) (data provided by the institution or review by external panel)	13	YES	16%
Quality of organization	Credit transfer system adopted by the institution (data provided by the institution)	9	NO	-
	Operations performable online (checklist: subscription, following lectures, examination, vote registration, ...) (data provided by the institution)	9	NO	-
	Percentage of student complaints or appeals solved/closed (data provided by the institution)	10	YES	24%
	Average time (days) for processing complaints/appeals (data provided by the institution)	5	NO	-
	Number of full-time equivalents (FTEs) employed for non-instructional, non-technical support services (providing assistance for admission, financial issues, registration, enrolment, etc.) weighted by student satisfaction for the service (data provided by the institution + student survey)	12	YES	29%
	Student satisfaction for room, laboratory and library facilities (through student survey)	10	YES	24%
	Student paperwork / online forms / front office time burden per week (through student survey)	4	NO	-
	Academic staff paperwork / online forms / front office / commissions time burden per week (through teacher survey)	8	NO	-
	Student satisfaction for organization (through student survey)	10	YES	24%
Sustainability of the Institution	Availability of an Institutional Strategic Plan for Online Learning (online vision statement, online mission statement, online learning goals and action steps, ...) (data provided by the institution)	17	YES	47%
	Percentage of curriculum changes resulting from an assessment of student learning (either formal or informal) within a fiscal year [a measure on increased flexibility within the curriculum	10	YES	28%

	development process to better respond to a rapidly changing world] (data provided by the institution)			
	Percentage of total institutional expenditure dedicated to online programmes (data provided by the institution)	9	YES	25%
	Ratio new students added / lost students per program (in the past year) (data provided by the institution)	6	NO	-
	Ratio new students added in the past year / students still in the program (data provided by the institution)	6	NO	-
	Percentage of Open Educational Resources used on the total of learning materials (data provided by the institution)	5	NO	-
	Percentage of Massive Open Online Courses on the total number of courses offered (data provided by the institution)	1	NO	-
Quality of the technological infrastructure	Student satisfaction with the overall learning platform (through student survey)	14	YES	39%
	Composite measure of server error rates, average response times, peak response times, and uptime (through technical institutional survey)	8	NO	-
	Checklist of functionalities supported by the platform (through technical institutional survey)	5	NO	-
	Measure of data security self-assessment (through technical institutional survey).	7	NO	-
	Measure of compliance with the accessibility guidelines WCAG 2.0 (through technical institutional survey)	13	YES	36%
	Availability of a plan for system maintenance and contingency management (data provided by the institution)	8	NO	-
	Measure of interoperability (Interoperability with external open sites (e.g., social media, DropBox, Google Drive), interoperability between LMSs (Learning Management Systems), information and teaching/learning materials exchange (LTI, SCORM, ...), Single sign-on (SSO) access control, etc. (data provided by the institution)	9	YES	25%

* indicator retained due to ties

Examining the results, we can see that indicators' weight significantly vary, with some indicators selected far more than others (e.g., in the criterion 'Reputation/impact' weights range from 8% to 19%). This aspect suggests that weights should not be disregarded and that some indicators may, indeed, be more important than others.

3 Final list of CODUR criteria and indicators

Table 3 reports the final list of CODUR criteria and indicators for evaluating online Higher Education Institutions, as it resulted from the Delphi Study.

Table 3 – Final list of CODUR criteria and observable indicators and weight.

Criterion	Observable indicator	Weight
Quality of teaching & learning	Student satisfaction of the overall learning experience (through student survey)	16%
	Student satisfaction regarding adequacy of the adopted pedagogical approaches to the learning objectives (through student survey)	16%
	Institutional support for learning design (in terms of tools, formats, etc.) (data provided by the institution)	15%
	Percentage of courses that propose personalized paths to reach the learning objectives (for example offering different materials/activities depending on culture, learning style, background, etc.) (data provided by the institution or review by external panel)	11%
	Student satisfaction regarding learning materials (through student survey)	15%
	Percentage of courses/examinations that make use of diverse forms of assessment (quantitative and qualitative approaches, human-based and technology-based tools, etc.) (data provided by the institution or review by external panel)	14%
	Student and teacher satisfaction regarding performance reports (through student and teacher survey)	11%
Quality of student support	Student satisfaction regarding interactions with teachers/tutors (through student survey)	56%
	Student satisfaction with technology support (including Helpdesk, FAQ, wizards, support material and initial training) (through student survey)	44%
Quality of teacher support	Teacher/tutor satisfaction with technology support (including Help desk, FAQ, wizards, support material and initial training) (through teacher survey)	35%
	Number of hours of training devoted to teaching staff concerning online learning per year (data provided by the institution)	22%
	Teacher/tutor satisfaction of training opportunities (through teacher survey)	24%
	Teacher/tutor satisfaction with feedback on their courses derived from students' surveys (through teacher survey)	18%

Reputation/Impact	Percentage of credits given in service-learning activities, in relation to total number of credits. Service Learning involves students in community service activities and applies the experience to personal and academic development. Service-learning takes place outside the HEI (data provided by the institution)	11%
	Number of clicks/likes/shares/comments/followers/impressions on academic social networks, such as Academia.edu, ResearchGate etc. (data provided by the institution)	8%
	Percentage of post-graduated actively engaged after graduation (data provided by the institution)	11%
	Percentage of former students employed in job sectors matching their degree (data provided by the institution)	19%
	A composite measure taking into account the existence of joint/dual degree programmes, the inclusion of study periods abroad, the % of international (degree and exchange) students, the % of international academic staff (data provided by the institution)	16%
	The number of student internships (total / per year) (data provided by the institution)	8%
	The number of student mobility (total / per year) (data provided by the institution)	14%
	The proportion of external research revenues - apart from government or local authority core/recurrent grants – that comes from regional sources (i.e. industry, private organisations, charities) (data provided by the institution)	12%
Quality of research	Internal budget devoted to research on online learning and teaching per Full Time Equivalent (FTE) academic staff (data provided by the institution)	16%
	Percentage of Full Time Equivalent (FTE) staff involved in research on online learning and teaching (data provided by the institution)	18%
	Yearly average n. of publications on online teaching & learning per Full Time Equivalent (FTE) academic staff (WoS or Scopus publications) (data provided by the institution or review by external panel)	18%
	Yearly average number of publications with authors from other countries per Full Time Equivalent (FTE) academic staff (WoS or Scopus publications) (data provided by the institution or review by external panel)	10%
	Internal budget devoted to disciplinary research per Full Time Equivalent (FTE) academic staff	10%
	External research income concerning disciplinary projects per Full Time Equivalent (FTE) academic staff	11%

	Yearly average n. of publications per Full Time Equivalent (FTE) academic staff (WoS or Scopus publications) (data provided by the institution or review by external panel)	16%
Quality of organization	Percentage of student complaints or appeals solved/closed (data provided by the institution)	24%
	Number of full-time equivalents (FTEs) employed for non-instructional, non-technical support services (providing assistance for admission, financial issues, registration, enrolment, etc.) weighted by student satisfaction for the service (data provided by the institution + student survey)	29%
	Student satisfaction for room, laboratory and library facilities (through student survey)	24%
	Student satisfaction for organization (through student survey)	24%
Sustainability of the Institution	Availability of an Institutional Strategic Plan for Online Learning (online vision statement, online mission statement, online learning goals and action steps, ...) (data provided by the institution)	47%
	Percentage of curriculum changes resulting from an assessment of student learning (either formal or informal) within a fiscal year [a measure on increased flexibility within the curriculum development process to better respond to a rapidly changing world] (data provided by the institution)	28%
	Percentage of total institutional expenditure dedicated to online programmes (data provided by the institution)	25%
Quality of the technological infrastructure	Student satisfaction with the overall learning platform (through student survey)	39%
	Measure of compliance with the accessibility guidelines WCAG 2.0 (through technical institutional survey)	36%
	Measure of interoperability (Interoperability with external open sites (e.g., social media, DropBox, Google Drive), interoperability between LMSs (Learning Management Systems), information and teaching/learning materials exchange (LTI, SCORM, ...), Single sign-on (SSO) access control, etc. (data provided by the institution)	25%

4 Design of a Toolbox for data gathering

In this section, the workflow for the design and the implementation of the Toolbox for data gathering is described.

Within the CODUR project, the design and test of the Toolbox for data gathering are carried out in the following phases (see Figure 2).

- Requirements' definition;
- Design and set up of the first version of Toolbox for data gathering;
- Development of an instrument for Toolbox validation and testing;
- Toolbox validation and testing with some non-paid partners (supporting institutions);
- Analysis of results.

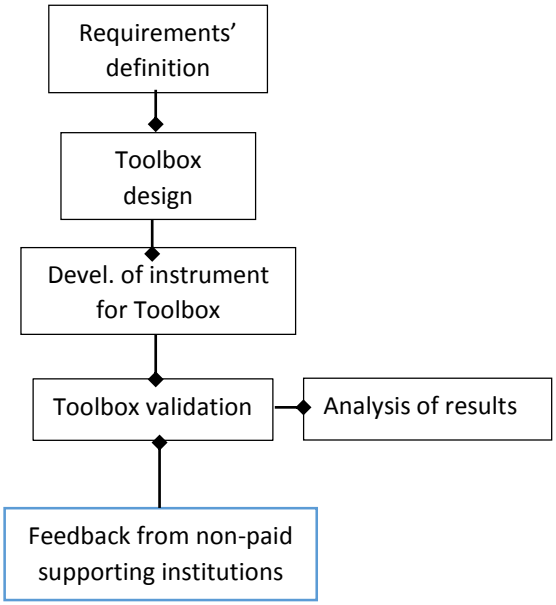

Figure 2 – Design and set up of the CODUR Toolbox. Workflow.

4.1 Requirements' definition

The Toolbox is a tool intended to support data collection (and management) from the online universities to be ranked. Starting from the results obtained in the previous phases of the CODUR project, we proceeded to perform an analysis of the Toolbox requirements, aimed at defining its main functions and capabilities at a goal level (Alexander & Beus-Dukic, 2009). In particular, the requirements' definition of the CODUR Toolbox revealed that the data to be collected belong to two different dimensions:

- desciptive data that need to be be collected from existing data sources (typically, institutional sources internal to the HEI);
- performance metrics that will be used as measures for comparing Online Universities and, in the final analysis, generating rankings.

Data collection (and management) is the core function of the CODUR Toolbox. Following the CIPO-model (i.e. Context-Input-Process-Output) (Scheerens, 1990), data belong to a context (in the CODUR project, the specific environment of an Online University) and represent a process (and its influences) at three different levels:

- Input, which defines the material, service and/or information that are used by the process in order to produce the outcome (output) (i.e. the logistical, human, and financial resources used by an Online University);
- Process, which is a defined sequence of activities. A process consists of procedural aspects (i.e. how an Online University works, how it uses resources, how it manages the inputs) and usually adds value to inputs in order to produce outputs for the service users (i.e. students or teachers);
- Output, which is the outcome of the process that is valuable to the service users (i.e. students or teachers) and concerns the actual achievements or products of an Online University.

It is important to note that, for the purposes of this work, the outputs of the process are those that must be measured, in terms of performance, to generate rankings. In fact, procedural aspects have more to do with accreditation of an Online University, and therefore deal with quality (in terms of assessment and assurance) by providing information and judgments (not ranking). This is strictly related to the value generated by the Online University as perceived by the service user (see Figure 3).

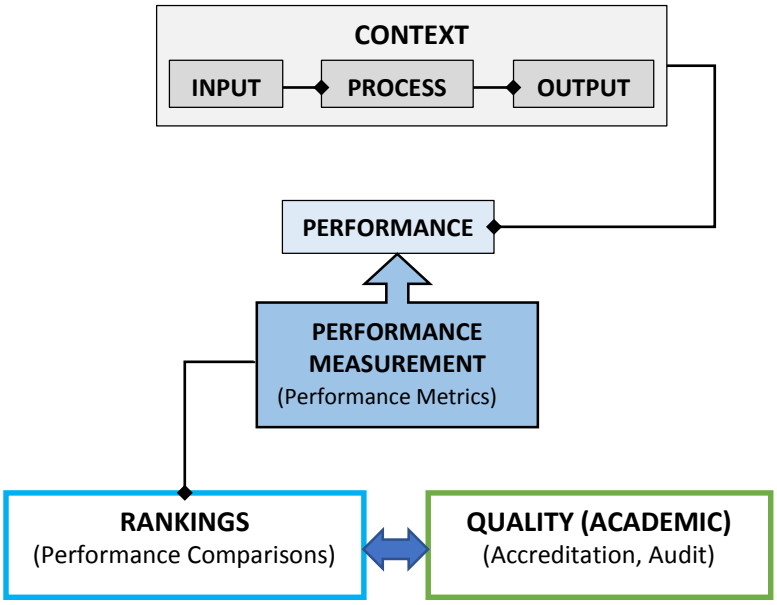

Figure 3 – Diagram: from performance indicators to rankings.

In the CODUR project, we are aware of the need to keep accreditation[3] (which evaluates the process) separated from ranking, where a synthetic score should be given based on objective data related to the outcome. The CODUR Toolbox is designed starting from this theoretical assumption, and its operative tools go exactly in the direction of collecting data related to the outputs.

[3] Accreditation agencies are objective, third-party institutions that follow a specific procedure for evaluating the quality of a university. If a university meets all of the standards required by the accreditation body, it gets its seal of approval. In some countries, accreditation is a prerequisite to issuance of a valid degree.
(https://www.accessmasterstour.com/articles/view/business-school-rankings-and-accreditations-how-to-use-them)

Generally speaking, data collection tools might be qualitative and/or quantitative. The CODUR Toolbox is based on surveys and rating scales as data collection tools, in order to elicit information about qualitative and quantitative attributes of Online Universities.

In reference to the CODUR Toolbox, data source is any existing person, group, or organization within an Online Institution that may provide information on whether the intended output (or outcome) occurred, and to what extent. In that sense, data source for the CODUR Toolbox are the following: students, teachers, institution (as a whole or at the departmental level), external reviewers (i.e. a panel of peer reviewers). Data type in the CODUR Toolbox includes alphanumeric strings such as numbers, percentages, composite scales, dichotomous scales (Yes/No), and Likert scales.

As stated previously, for the sake of the CODUR project, we are mostly interested in measuring and comparing data collected in reference to the outcomes. The measurement of outcomes is defined in terms of performance measurement and is obtained by means of synthetic and objective scores (performance metrics) for comparing Online Universities and generating rankings.

For this purpose, operational variables referring to specific empirically measurable features of Online Universities on which evidence can be collected have been identified, defined, and validated in terms of weighted indicators through the Delphi Study (see 2.1.3).

Performance metrics, at this point, have to be scaled so that, before applying the weightings, they have the same relative importance. This is necessary because some indicators (e.g., internal budget devoted to disciplinary research per FTE academic staff) are measured on a different order of magnitude than others (e.g., percentage of courses that propose personalized paths to reach the learning objectives). The former category of indicators risks to outweight the ranking, unless corrective scaling is applied before weighting indicators. One way to perform such a scaling is transforming each measure in the corresponding percentile rank in the distribution of evaluated HEIs. This kind of transformation has the single drawback that it makes measures sample-dependent, i.e. measures can only be interpreted relative to each other, and not as absolute measures. However, since the goal of the toolbox is *ranking*, rather than evaluation, it is possible to adopt this approach without losing goal-relevant information. Therefore, the collected data can be exploited to generate performance comparisons, or rankings, and to sort Online Universities into “ranking order”.

In such a scenario, rankings could be used for instance by service users to get a more precise idea about the value of an Online University, in comparison with others, with regard to a certain dimension.

4.2 Design and development of the Toolbox

The CODUR Toolbox is a digital tool for data collection, which can be shared with the identified stakeholders to collect information and data.

The Toolbox looks like a spreadsheet, structurally organized in 8 worksheets (i.e. “tabs”) which allow the user to easily switch from one criterion to another. Each tab corresponds to one of the eight criteria previously identified (see Table 1):

- Tab: “Teaching & Learning”;
- Tab: “Student Support”;
- Tab: “Teacher Support”

- Tab: “Reputation-Impact”;
- Tab: “Research”;
- Tab: “Organization”;
- Tab: “Sustainability”;
- Tab: “Technological Infrastructure”.

Each tab is divided in 6 columns and contains:

- [Column 1] The criterion.
- [Column 2] The list of the related observable indicators. The indicators are listed in successive rows.
- [Column 3] The weights of the observable indicators.
- [Column 4] The data sources to be used.
- [Column 5] The data types.
- [Column 6] The data inputs.

Content of cells from Colum 1 to Column 5 is blocked: this means that their values cannot be edited by the user, but only read. Conversely, content of cells of Colum 6 must be inputted by the user, who is required to provide the data input for the specific observable indicator.

See Figure 4 for taking a closer look to a mock-up of the “Teaching & Learning” tab.

Criterion	Observable indicator	Weight	Data source	Data type	Data input
Quality of teaching & learning	Student satisfaction of the overall learning experience	16,00%	Student survey	Likert scale	
	Student satisfaction regarding adequacy of the adopted pedagogical approaches to the learning objectives	16,00%	Student survey	Likert scale	
	Institutional support for learning design (in terms of tools, formats, etc.)	15,00%	Institution	Yes/No	
	Percentage of courses that propose personalized paths to reach the learning objectives (for example offering different materials/activities depending on culture, learning style, background, etc.)	11,00%	Institution	Percentage	
	Student satisfaction regarding learning materials	15,00%	Student survey	Likert scale	
	Percentage of courses/examinations that make use of diverse forms of assessment (quantitative and qualitative approaches, human-based and technology-based tools, etc.)	14,00%	Institution	Percentage	
	Student and teacher satisfaction regarding performance reports	11,00%	Student and teacher survey	Likert scale	

Figure 4 – The CODUR Toolbox. “Quality of teaching & learning” tab.

5 Toolbox validation and testing

The CODUR Toolbox has been tested with some non-paid supporting institutions. The main aim of the process of Toolbox validation and testing is to check whether and to what extent the data envisaged by the CODUR set of criteria and indicators can actually be provided by online Higher Education Institutions. This is because there are a number of issues that remain open and may be challenging in data collection and management, in the context of Online Universities. These include:

- Estimated time required for data collection. How long would it take to collect the data for an HE Institution? If the estimated time turns out to be too long, online Universities might fail to provide the data in a timely manner.
- Scheduling. Does the data need to be collected at a certain time or period of the year? Online Universities have internal and external constraints as far as time scheduling is concerned. Some data can be available at certain time of the year (for example, the number of new enrollments in one year might be available only after the enrollment phase is officially closed, etc.) and this might impose limitations on data collection schedule.
- Data availability. Has this data already been collected / is this data (already) available? Most universities collect data about their activities for internal formative assessment purposes. If these data can be made public, they are obviously more readily made available by the HEI with virtually no cost.
- Publicity/Privacy Issue. Can this data be made public / open? Online Universities might have constraints in terms of data openness and this needs to be taken into consideration to estimate if a datum will be realistically available or not.

In order to understand whether and to what extent these issues might become an actual barrier that prevents (some of the) data to be collected, an ad hoc data collection instrument has been implemented by CNR-ITD. The instrument for Toolbox validation has been structured as follows (see Table 4):

- overall, the instrument structure is similar to the Toolbox itself;
- a number of additional columns have been included in the instrument, that contain specific questions, each one addressing one of the above mentioned issues, namely:
    - [Column 6] The estimated time required for data collection, in the form of the following question: “How long would it take to collect the data?”
    - [Column 7] The scheduling for data collection, in the form of the following question: “Does the data need to be collected at a certain time or period of the year?”
    - [Column 8] The estimated time required for data collection, in the form of the following question: “Has this data already been collected / is this data (already) available?”
    - [Column 9] The publicity of the data, in the form of the following question: “Can this data be made public / open?”
    - [Column 10] In this field respondents should provide “Data example” that are aimed in this testing phase not so much at measuring the actual situation, but rather at providing an example of the kinds of data we could possibly expect.

This way, feedback collected in columns 6, 7, 8, and 9 will be useful to validate the Toolbox. Furthermore, feedback collected in column 10 will help to collect, with reference to the proposed model, data useful for:

- Designing a basic data repository (i.e. pivot table) with filtering tools;
- Implementing some preliminary data representation and visualization (such as league tables, charts or graphics).

To support the non-paid institutions in the test, a compilation guide has also been made available (Annex 2), along with a video tutorial, aimed to introduce the concept of "ranking". The video is available at:
https://www.youtube.com/watch?v=KCtnD2WJS48&feature=youtu.be

The data collected during the validation are available on the project GoogleDrive: https://drive.google.com/drive/folders/1AS_Dixl_0M22T3ueCyLERx-BAKhnv8st

The results of the data analysis are contained in Section 6, along with a specific discussion on the testing phase and the toolbox itself (Section 6.3).

Table 4 – The instrument for Toolbox validation and testing. “Quality of teaching & learning” tab.

[1] Criterion	[2] Observable indicator	[3] Weight	[4] Data source	[5] Data type	[6] How long would it take to collect this data?	[7] Does this data need to be collected at a certain time of the year?	[8] Has this data already been collected / is this data already available?	[9] Can this data be made public/open?	[10] Data example
Quality of teaching & learning	Student satisfaction of the overall learning experience	16,00%	Student survey	Likert scale					
	Student satisfaction regarding adequacy of the adopted pedagogical approaches to the learning objectives	16,00%	Student survey	Likert scale					
	Institutional support for learning design (in terms of tools, formats, etc.)	15,00%	Institution	Yes/No					
	Percentage of courses that propose personalized paths to reach the learning objectives (for example offering different materials/activities depending on culture, learning style, background, etc.)	11,00%	Institution	Percentage					
	Student satisfaction regarding learning materials	15,00%	Student survey	Likert scale					
	Percentage of courses/examinations that make use of diverse forms of assessment (quantitative and qualitative approaches, human-based and technology-based tools, etc.)	14,00%	Institution	Percentage					
	Student and teacher satisfaction regarding performance reports	11,00%	Student and teacher survey	Likert scale					

6 Discussion and Conclusions

The main aim of the CODUR project was to devise and test a set of criteria and indicators specifically devoted to the evaluation and ranking of online Higher Education Institutors. At end of this long and complex process, there are a number of considerations and reflections that can be made, especially regarding: 1) the process which led to the definition of the CODUR criteria and indicators; 2) the proposed criteria and indicators themselves; 3) the testing phase of the indicators and the toolbox conducted with the non-paid Universities.

Below, we propose the main lessons learnt, along with some recommendations deriving from the overall experience.

6.1 Discussion on the process

The identification of criteria and indicators started at the beginning of the project with an analysis regarding the state of art of online education in the world (Giardina, Guitert, & Sangrà, 2017) and a parallel study on the existing ranking systems (Brasher et al., 2017).

The former analysis pointed out that online education is a global trend and it is increasingly growing (even if with different levels of expansion and despite a certain difficulty to retrieve consistent data); this confirms in any case the need to have tools able to evaluate and compare the quality of the services offered. At the same time, the latter study pointed out that – despite the high number of available ranking systems – we are far from having a unique or standardized way to measure quality and rank universities. On the contrary, the available systems are very heterogeneous in nature and the indicators adopted are very different and not always understandable and/or transparent. This made our work, that needed to build on already existing systems, quite challenging and difficult. Given that in our project U-Multirank had been already identified as a possible candidate for integrating the CODUR criteria and indicators, we devoted particular attention to its analysis, in an effort to create a set of criteria and indicators somehow 'compatible' with the ones already proposed within U-Multirank.

In line with the CODUR proposal, the project took a participatory approach to the definition of the criteria and indicators, i.e. the design phase was not something that happened within the project boundaries, but, on the contrary, it involved several stakeholders and informants and the broader HE community. This was done in the assumption that taking into consideration the points of view of all the relevant stakeholders (both individual people and bodies) is crucial for the criteria and indicators to be later on recognized, accepted and ultimately used. Furthermore, this should lead to a more exhaustive set of criteria and indicators, able to capture and evaluate all the aspects and variants at play. In particular, the main approach adopted to define the set of criteria and indicators was the Delphi Study. This allowed us to involve a number of international experts (40 at Round 1 and 21 at Round 2 respectively) and consult them. Such participatory approach, able to engage people with solid expertise in the field, but at the same time providing different perspectives and points of view, is for sure one of the main added value of the CODUR project. In particular, it is important to stress the fact that – for the first time at least to our knowledge – *learners* have been involved in the process of defining the criteria, in such a way that the point of view of the 'final user' of online Institutions can be taken into due consideration. This was far from being easy, because in many institutions, involving learners in surveys requires to set up procedures to assure compliance with ethical and privacy regulations.

Another challenge we faced had to do with the use of terminology that is not always uniquely defined, and the online interactions with the Delphi experts could lead to misinterpretations. Providing definitions for terms at the beginning of each survey item was the best solution we found to mitigate the risk of misunderstandings, but of course, we cannot guarantee some of the terms have not been misinterpreted by people while answering. In this sense the occasion provided by the EMEM workshop (Pozzi et al., 2017) and the opportunity we had to discuss the same items with experts during a face-to-face event, encouraged us, given that the results of the workshop reflect and confirm the outcomes of the Delphi (Round I; see Pozzi et al., 2017). Moreover, the workshop was also an occasion to go more in depth into the discussion, by adding qualitative information to the Delphi, which – even if envisaged the possibility to provide textual comments – was far more profitable at the level of quantitative information.

So conceived, the process, along with the actions put in place by the consortium, turned out to be quite effective in terms of inputs and feedback collected.

6.2 Discussion on the CODUR criteria and indicators

As a result of the process conducted during the project, we came out with a set of 8 criteria and a total of 38 observable indicators (on average: 4.75 indicators per criterion). Actually, one of the most challenging and often questioned aspect of rankings is their ability to capture and measure the complexity of reality with a reasonable number of indicators. While some ranking systems are criticised for their being reductionist, others are for their complexity. Even U-Multirank, which was an attempt to propose a manageable number of transparent and ease-to-read indicators, has been recommended to scale down and simplify its indicators (Wächter, Kelo, Lam, Effertz, Jost & Kottowski, 2015). Therefore, keeping in mind this recommendation, in CODUR we have tried to keep the number of indicators as low as possible. To do so, we examined the indicators already present in U-Multirank and kept those that could be applied to online institutions, in such a way to avoid adding new ones, if we felt the already existing ones were fitting – to some extent - with our purposes.

Besides, following the recommendations given by Wächter et al. (2015, pg. 78), i. e. to conduct research especially on “adequate and internationally comparable indicators for the quality of teaching”, which seems missing or unsatisfactory in most existing ranking systems, we devoted particular attention to the area of Teaching and Learning, by conceiving an *ad hoc* criterion and then defining a number of specific indicators, which should – according to the experts involved in the Delphi – evaluate such a crucial dimension. Interestingly, the observable indicators defined for this criterion, put an emphasis on the pedagogical approaches adopted, the learning design phase, the personalization opportunities and the kinds of assessment available. These observable indicators seem particularly reasonable and contrast with some of the indicators adopted by other existing ranking systems (such as for example U-Multirank), which focus on indicators of outcome (such as the percentages of graduates on time, the academic staff with doctorates, etc.), or on other aspects (such as for example the library or the laboratory facilities), which have to do more with organizational and infrastructural aspects. Overall, we think the observable indicators put forward by our Delphi Study for the Teaching and Learning criterion are so significant that we recommend them to be considered for inclusion also in other existing ranking systems, addressing online or traditional Universities.

As far as the indicators identified for the “Reputation /Impact”, “Quality of research”, “Quality of organization”, and “Sustainability of the institution”, most of them come from U-Multirank and can be applied to both online and traditional HE institutions.

On the contrary, the indicators that have to do with the “Student support”, “Teacher support” and “Technological infrastructure”, target aspects that are far more important for online Institutions, rather than for traditional ones. Despite this, we believe these dimensions should be considered nowadays also by face-to-face institutions, as in any case blended approaches are becoming increasingly spread even in traditional contexts.

Furthermore, to address the issue of transparency of indicators, which is often questioned for existing ranking systems, we have provided indicators along with their weighting, in such a way as to make it explicit to what extent each aspect is important in relation to the dimension under the lens. To be noted that weightings, too, were assigned on the basis of opinions expressed by the experts in the Delphi.

Observable indicators have been kept as ‘simple’, ‘operative’ and ‘raw’ as possible and we have tried to avoid complex or aggregated indicators, so as to support readability and ease of use for the final user. This does not mean that some of them could be subsequently aggregated, to provide a more concise view on data.

6.3 Discussion on the testing phase and the toolbox

The instrument for the validation of the CODUR project toolbox was tested by a suite of online and blended higher education institutions. The aim of this test, as it has been referred above, was to document whether online higher education institutions would be able to easily provide the information required by each of the CODUR indicators.

The analysis of the indicators referring to the quality of teaching and learning reveals a promising scenario. Most of the indicators were already available and used by the participant universities or could be easily introduced in the near future. More concretely, universities periodically collected data on student satisfaction of the overall learning experience, student satisfaction regarding adequacy of the adopted pedagogical approach to the learning objectives, and student satisfaction regarding learning materials. Since these indicators are central for the quality of the institution, they were collected quarterly. These central indicators could be complemented with another two indicators identified by experts during the previous phases of the CODUR project: the institutional support for the learning design and student and teacher satisfaction regarding performance reports. The data for these indicators was either already available for universities or could be obtained in the next two or more years. Two more indicators were deemed less useful by institutions: the ‘percentage of courses that propose personalized paths to reach the learning objectives’ and the ‘percentage of courses/examinations that make use of diverse forms of assessment’. Data regarding the quality of teaching and learning was mostly available to the universities through student satisfaction surveys and were collected annually or biannually.

Another relevant criteria for the assessment of online education is the quality of the student support offered by an online or blended higher education institution. The participant higher education institutions reported collecting data on student satisfaction regarding interactions with teachers/tutors. This data was collected quarterly, using student satisfaction surveys that include Likert scale questions on the topic. Half of these institutions were also taking into account student satisfaction with technology support. The other half, however, considered this indicator too difficult to manage. Data on this indicator was also collected with student satisfaction surveys.

Quality of teacher support was assessed with a plurality of indicators. Most of the participating institutions counted with data on the teachers and tutors satisfaction with technology support and on their satisfaction with training opportunities. In contrast, data on the 'number of hours of training devoted to teaching staff concerning online learning per year' and 'teacher/tutor satisfaction with feedback on their courses derived from students' surveys' were considered more difficult to manage. Some universities counted with these indicators or could make use of them in the future and other considered that they were just too hard to handle. The difficulty with the indicator 'teacher/tutor satisfaction with feedback on their courses derived from students' surveys' is that it requires two steps: surveying students and then surveying teachers about it. Universities proved to be more comfortable with data that could be obtained in one single step. As online learning becomes more pervasive and quality assessment improves, more universities may be able to consider this indicator. Data on the quality of teacher support is currently mostly gathered annually or biannually with teacher surveys and/or staff climate surveys.

A further criterion to be considered is the reputation and impact of online and blended institutions. The analysis of this criterion provides encouraging results. Most of the participant higher education institutions already gathered data on and used five of the eight suggested indicators. Two of these indicators refer to the success of students after graduation. More concretely, almost all institutions counted with data on the percentage of post-graduated actively engaged after graduation and on the percentage of former students employed in job sectors matching their degree. Another frequently used indicator refers to the availability of external research revenues, besides government funds, that come from regional sources. All institutions had data on this topic or were hoping to introduce it in the near future. Likewise, most institutions could report on the percentage of international (degree and exchange) students at their institution and on the number of student mobility. Data on impact and reputation was mostly acquired quarterly or annually from institutional sources.

Some indicators on reputation and impact were more problematic. First, most of the participant universities considered the indicator 'Percentage of credits given in service-learning activities, in relation to the total number of credits' too hard to manage. This is due to the fact that there is not a common strategy for online institutions to offer service-learning activities and these can take place in a variety of ways, for instance through online simulation or through practices at the student workplace. Furthermore, some national agencies on higher education consider that institutions that demand students to complete service-learning activities offline cannot be considered completely online institutions regardless on how the rest of the instruction takes place.

Second, the indicator 'number of clicks/likes/shares/comments/followers/impressions on academic social networks', beyond the number of viewers of the institution's homepage, was also considered very difficult to manage. Advances in data mining and big data analysis may encourage universities to collect more information on this criterion.

Three further problematic indicators refer to joint/dual degree programmes, the inclusion of study periods abroad and the percentage of international academic staff. The participant institutions provided a wide range of considerations regarding these indicators, from specific information on how they effectively gather this information to their conviction on the impossibility of their institution of reporting on the indicators. Mixed reports are also frequent.

These diversity of responses and the common lack of data may be due to the fact that it is not a priority for online and blended higher education institution to determine the location and provenance of their students and faculty as most of the academic activity takes place online. Finally, all institutions reported their interest of incorporating an indicator regarding the number of student internships.

The criterion that offers more room for the improvement of universities is that of the quality of research. Indeed, data in this area was very scarce and only few universities indicated having data for some specific indicators. Nevertheless, the interest for incorporating indicators in this area was general. Universities suggested different incorporation times ranging from three months to more than two years. These differences respond solely to the characteristics of each university. The only indicator that was considered problematic out of the seven suggested by the CODUR team was 'the yearly average number of publications with authors from other countries per full time equivalent academic staff'. Again, this suggests that knowing the provenance or location of their staff may not be as relevant for online providers as it had been for their face-to-face counterparts.

The landscape regarding indicators on the quality of the organization is more varied. Most institutions reported having indicators on the percentage of student complains or appeals solved/closed and on student satisfaction for room, laboratory and library facilities. Data for these indicators was collected annually or biannually using institutional sources and student surveys. The participating universities provided very contrasting responses regarding another indicator in this area: the student satisfaction for the organization'. Half of the organizations counted with this data while the other half did not and could not even consider incorporating the indicator in the future. None of the institutions counted with data regarding the number of full-time equivalents employed for non-instructional, non-technical support services (providing assistance for admission, financial issues, registration, enrolment, etc.) weighted by student. However, most of the universities expressed their interest for incorporating this indicator in the future.

Data on the sustainability of the institution and on the quality of the technological infrastructure is easier to report on. In fact, these criteria refer to the everyday functioning of the institution and most indicators were already available at the participating institutions. Regarding the sustainability of the institution, most participant universities counted with data on the availability of an institutional strategic plan for online learning and on the percentage of total institutional expenditure dedicated to online programmes. In contrast, universities presented very different responses regarding the indicator 'percentage of curriculum changes resulting from an assessment of student learning (either formal or informal) within a fiscal year'. The variety of responses may be due to the fact that the gathering of data for this indicator requires two steps instead of one and this, as mentioned above, discourages universities from collecting that data.

Finally, all participant universities reported being able of measuring the quality of their technological infrastructure through their students' satisfaction with the overall learning platform. They also all expressed the interest of including an indicator regarding the compliance of their institutions with the accessibility guidelines WCAG 2.0. Some universities counted as well with a measure of interoperability gathered through subjective evaluations or automated performance checks.

Overall, the test of the CODUR Toolbox reveals that it is a very useful instrument for online and blended higher education institutions to gather and manage data concerning their performance and outcomes. Likewise, it shows the utility of most of the CODUR criteria and indicators on the quality of online education provision for online and blended universities. A slight difficulty has been detected and it is that of determining the origin and current location of professors, researchers, and students since online everyday teaching and learning allows for an internationalization of the faculty and student body. Thus, practice reveals that perhaps this element may not be as essential to evaluate the quality of online and blended higher education institutions as first considered by experts.

7 References

Alexander, I. F., & Beus-Dukic, L. (2009). *Discovering requirements: How to specify products and services*. Chichester, UK: John Wiley & Sons.

Brasher, A., Holmes, W. & Whitelock, D. (2017*). A means for systemic comparisons of current online education quality assurance tools and systems. CODUR Project Deliverable IO1.A2*.

Dalkey, N., & Helmer, O. (1963). An experimental application of the Delphi method to the use of experts. *Management Science*, *9*(3), 458-467. https://doi.org/10.1287/mnsc.9.3.458

European Commission (2014). *Report to the European Commission on new modes of learning and teaching in Higher Education*. Luxembourg, LUX: Publications Office of the European Union. https://doi.org/10.2766/81897

Giardina, F., Guitert, M., & Sangrà, A.. (2017*). A means for systemic comparisons of current online education quality assurance tools and systems. CODUR Project Deliverable IO1.A1*.

Pozzi, F., Manganello, F., Passarelli, M., & Persico, D. (2017). *The State of Art of Online Education Report. CODUR Project Deliverable IO1-A3*.

Rowe, G., & Wright, G. (1999). The Delphi technique as a forecasting tool: issues and analysis. *International Journal of Forecasting*, *15*(4), 353-375. https://doi.org/10.1016/S0169-2070(99)00018-7

Scheerens, J. (1990). School effectiveness research and the development of process indicators of school functioning. *School effectiveness and school improvement*, *1*(1), 61-80. https://doi.org/10.1080/0924345900010106

Van Vught, & Ziegele, F. (2011). *Design and testing the feasibility of a multidimensional global university ranking. Final Report*. CHERPA-Network. Retrieved from: https://www.eurashe.eu/library/quality-he/U-Multirank_Design%20and%20Testing%20the%20Feasibility%20of%20a%20Multidimensional%20Global%20University%20Ranking.pdf

Wächter, B., Kelo, M., Lam, Q. K. H., Effertz, P., Jost, C. & Kottowski, S. (2015). *University quality indicators: A critical assessment*. Directorate General for Internal Policies – Policy Department B: Structural and Cohesion Policies – Culture and Education. Retrieved from: http://www.europarl.europa.eu/RegData/etudes/STUD/2015/563377/IPOL_STU%282015%29563377_EN.pdf

Annex 1 - Questionnaire Round 2 of the Delphi study

CODUR - Towards the creation of an online dimension for University Rankings (Round 2)

Dear participant,
Thank you for your agreeing to participate to the second round of the CODUR Delphi Study.

The aim of this study is to identify ***indicators for evaluating the quality of online higher education institutions***. For the purpose of this study, we are defining "online higher education institutions" as those that offer degrees.

The first round of the Delphi Study focused on finding the features of online higher institutions that should be taken into account for evaluation. In this second round, we will focus on finding ***measurable indicators*** that can be used for ranking online higher education institutions.

In the first part of the survey, you will be presented with the results of the first round. Please read them carefully before proceeding.

In the second part of the survey, you will be presented with several ***measurable indicators*** for university raking systems, clustered in eight themes. For each theme, you will be asked to select what you believe to be the best indicators. We ask you to select ***no more than half*** of the indicators for each theme. The number of maximum selectable indicators will be indicated at the top and bottom of each theme's page.

The themes are:

- Teaching and Learning,
- Research,
- Organization of the institution,
- Sustainability of the institution,
- Reputation and impact of the institution,
- Student support,
- Teacher support,
- Technological infrastructure.

In the survey, these themes will be presented to you in a random order.

The information that you provide in this survey will be aggregated and ***anonymised*** before publication. We do, however, request that you provide a little information about your knowledge of this area and would be grateful for the opportunity to contact you if we have any queries or further questions in relation to your comments (this is entirely voluntary).

The CODUR partners
[Contact email: pozzi@itd.cnr.it]

There are 69 questions in this survey

Main results from round I of the Delphi Study

Output

Relative importance of criteria

In the first round of the Delphi Study, we asked participants to rank in order of importance nine criteria for evaluating online Higher Education Institutions.

The criteria, *from most important to least important*, are:

- Quality of the learning experience
- Quality of teaching
- Quality of student support
- Quality of the technological infrastructure
- Quality of teacher support
- Quality of research
- Quality of organization
- Sustainability of the institution
- Reputation of the institution

Quality of the learning experience and Quality of teaching have generally been considered the two most important criteria for evaluating online institutions. Following qualitative feedback, these criteria were combined into one (quality of teaching and learning) for the second round of the study.

Quality of research, Quality of teacher and student support, and Quality of the infrastructure are considered of medium importance; Quality of organization, Sustainability of the institution and Reputation form the lowest tier of indicators of quality.

Indicators' importance

For each of the nine preliminary CODUR criteria, participants were presented with their indicators and were asked to rate their importance for assessing the related criterion. Rating was performed on a 0-4 scale (from 0 = not at all important to 4 = extremely important).

Descriptive statistics for all ratings are reported in the full report (see link at the bottom of the page).

Examining the results, we can see that most indicators are deemed very important for assessing their criterion. The only indicators with a score below 2.5 - which could still be considered a high threshold - are the existence of newsletters for communicating with students and teaching staff (measuring quality of student and teacher support, respectively), support to alumni community building (student support), presence of decentralized structures on the territory (organization), and size of the institution (sustainability of the institution).

Moreover, there is moderate agreement between raters for indicator importance.

In the survey, participants were also allowed to propose additional indicators for each criterion, if they deemed it appropriate. The list of proposed indicators is available in the full report, along with additional details on the results and methodologies. If you are interested, the report is accessible at the following link:

https://drive.google.com/file/d/1GWFt0ZXJZBeC_o5el9sCO2qrHHPhs9r_/view?usp=sharing

Some questions about you
Gender

Please choose **only one** of the following:

- ○ Female
- ○ Male

Age
Only numbers may be entered in this field.

Please write your answer here:

Country

Please write your answer here:

You are...
Check all that apply

Please choose **all** that apply:

- ☐ ...a student
- ☐ ...a researcher
- ☐ ...an educator (higher education)
- ☐ ...an educator (school)
- ☐ ...an educator (workplace)
- ☐ ...a policy maker
- ☐ Other:

How well informed are you about University ranking systems?
Choose one of the following answers

Please choose **only one** of the following:

- ○ Not at all informed
- ○ Slightly informed
- ○ Well informed
- ○ Very well informed

Choose the best indicators for quality of teaching and learning
Please choose ***from 0 to 6*** (half) of the following indicators. The indicators you select are those that, in your opinion, are best suited to measure the ***quality of teaching and learning***. For each indicator, the data source has been reported in brackets. If you have any comments on the measurable indicators, you can provide them in the text box at the end of this page.

Quality of overall learning experience
Check all that apply

Please choose **all** that apply:

- ☐ Student satisfaction of the overall learning experience (through student survey)

Quality of pedagogy /methodology
Check all that apply

Please choose **all** that apply:

- ☐ Student satisfaction regarding adequacy of the adopted pedagogical approaches to the learning objectives (through student survey)

Quality of course / learning design
Check all that apply

Please choose **all** that apply:

- ☐ Institutional support for learning design (in terms of tools, formats, etc..) (data provided by the institution)
- ☐ Percentage of courses that propose personalized paths to reach the learning objectives (for example offering different materials/activities depending on culture, learning style, background, etc.) (data provided by the institution or review by external panel)
- ☐ Percentage of courses that support self-regulated learning (e.g. tools for deadline management, tools for progress tracking, self-evaluation tools) (data provided by the institution or review by external panel)

Quality of learning materials /activities
Check all that apply

Please choose **all** that apply:

- ☐ Student satisfaction regarding learning materials (through student survey)
- ☐ Student satisfaction regarding proposed activities (through student survey)

Quality of assessment
Check all that apply

Please choose **all** that apply:

- ☐ Percentage of courses/examinations that make use of diverse forms of assessment (quantitative and qualitative approaches, human-based and technology-based tools, etc.) (data provided by the institution or review by external panel)

Tracking of online interactions
Check all that apply

Please choose **all** that apply:

- ☐ Percentage of courses that provide performance reports to learners & teachers by means of learning analytics (data provided by the institution or review by external panel)
- ☐ Student and teacher satisfaction regarding performance reports (through student and teacher survey)

Standards for regulating teacher-student interactions
Check all that apply

Please choose **all** that apply:

- ☐ Existence of suggested standards for feedback provision (e.g. time threshold, ...) (data provided by the institution)

Quality control and teacher assessment
Check all that apply

Please choose **all** that apply:

- ☐ Frequency of course/programme evaluation (data provided by the institution)

Comments

Please write your answer here:

Before continuing, please make sure that you selected no more than *6* indicators.

Choose the best indicators for quality of student support
Please choose ***from 0 to 2*** (half) of the following indicators. The indicators you select are those that, in your opinion, are best suited to measure the ***quality of student support***. For each indicator, the data source has been reported in brackets. If you have any comments on the measurable indicators, you can provide them in the text box at the end of this page.

Quality of interactions between educators and students
Check all that apply

Please choose **all** that apply:

- ☐ Ratio tutors/students (data provided by the institution)
- ☐ Student satisfaction regarding interactions with teachers/tutors (through student survey)

Technology support
Check all that apply

Please choose **all** that apply:

- ☐ Student satisfaction with technology support (including Helpdesk, FAQ, wizards, support material and initial training) (through student survey)
- ☐ Helpdesk average response time to students (data provided by the institution)

Orientation services to help learners taking decisions about their learning path
Check all that apply

Please choose **all** that apply:

- ☐ Student satisfaction with orientation services (through student survey)

Comments

Please write your answer here:

Before continuing, please make sure that you selected no more than *2* indicators.

Choose the best indicators for quality of teacher support
Please choose ***from 0 to 4*** (half) of the following indicators. The indicators you select are those that, in your opinion, are best suited to measure the ***quality of teacher support***. For each indicator, the data source has been reported in brackets. If you have any comments on the measurable indicators, you can provide them in the text box at the end of this page.

Technology support
Check all that apply

Please choose **all** that apply:

- ☐ Teacher/tutor satisfaction with technology support (including Help desk, FAQ, wizards, support material and initial training) (through teacher survey)
- ☐ Helpdesk average response time to teachers/tutors (data provided by the institution)

Opportunities for teaching staff to be trained
Check all that apply

Please choose **all** that apply:

- ☐ Number of hours of training (or equivalent) made available for teachers/ tutors by the institution per year (data provided by the institution)
- ☐ Number of hours of training devoted to teaching staff concerning online learning per year (data provided by the institution)
- ☐ Number of hours of training devoted to teaching staff concerning code of ethics per year (data provided by the institution)
- ☐ Teacher/tutor satisfaction of training opportunities (through teacher survey)

Support to teaching staff
Check all that apply

Please choose **all** that apply:

- ☐ Teacher/tutor satisfaction with Community building tools made available by the institution (through teacher survey)
- ☐ Teacher/tutor satisfaction with design patterns and OER (Open Education Resources) repositories made available by the institution (through teacher survey)
- ☐ Teacher/tutor satisfaction with feedback on their courses derived from students' surveys (through teacher survey)

Comments

Please write your answer here:

Before continuing, please make sure that you selected no more than *4* indicators.

Choose the best indicators for reputation/impact
Please choose ***from 0 to 7*** (half) of the following indicators. The indicators you select are those that, in your opinion, are best suited to measure the ***reputation/impact of the institution***. For each indicator, the data source has been reported in brackets. If you have any comments on the measurable indicators, you can provide them in the text box at the end of this page.

Social impact & responsibility
Check all that apply

Please choose **all** that apply:

- ☐ Percentage of credits given in service-learning activities, in relation to total number of credits. Service Learning involves students in community service activities and applies the experience to personal and academic development. Service-learning takes place outside the HEI (data provided by the institution)

Communication strategies & Visibility on academic social networks
Check all that apply

Please choose **all** that apply:

- ☐ SEO (Search Engine Optimization) on institutional Websites (data provided by the institution). SEOs are strategies and activities aimed to improve visibility of a website on Internet search engines.
- ☐ Position on Webometrics University Ranking (data provided by the institution). The Webometrics Ranking of World Universities, also known as Ranking Web of Universities, is a ranking system for the world's universities.
- ☐ Number of clicks/likes/shares/comments/followers/impressions on academic social networks, such as Academia.edu, ResearchGate etc. (data provided by the institution)

Job opportunities for graduates
Check all that apply

Please choose **all** that apply:

- ☐ Percentage of post-graduated actively engaged after graduation (data provided by the institution)
- ☐ Percentage of former students employed in job sectors matching their degree (data provided by the institution)

International orientation
Check all that apply

Please choose **all** that apply:

- ☐ A composite measure taking into account the existence of joint/dual degree programmes, the inclusion of study periods abroad, the % of international (degree and exchange) students, the % of international academic staff (data provided by the institution)

Internship and mobility opportunities
Check all that apply

Please choose **all** that apply:

- ☐ The number of student internships (total / per year) (data provided by the institution)
- ☐ The number of student mobility (total / per year) (data provided by the institution)

Relationship with the territory
Check all that apply

Please choose **all** that apply:

- ☐ Percentage of student internships in the region (data provided by the institution)
- ☐ The number of theses (BA and MA) with regional organisations (data provided by the institution)
- ☐ The proportion of external research revenues - apart from government or local authority core/recurrent grants – that comes from regional sources (i.e. industry, private organisations, charities) (data provided by the institution)
- ☐ The percentage of graduates who found their first job (after graduation) in the region where the university is located (data provided by the institution)

Representation on national forums
Check all that apply

Please choose **all** that apply:

- ☐ The number of national forums joined (data provided by the institution)

Comments

Please write your answer here:

Before continuing, please make sure that you selected no more than *7* indicators.

Choose the best indicators for quality of research
Please choose ***from 0 to 6*** (half) of the following indicators. The indicators you select are those that, in your opinion, are best suited to measure the ***quality of research***. For each indicator, the data source has been reported in brackets. If you have any comments on the measurable indicators, you can provide them in the text box at the end of this page.

Research in online teaching & learning (research groups, research projects, etc.)
Check all that apply

Please choose **all** that apply:

- ☐ Internal budget devoted to research on online learning and teaching per Full Time Equivalent (FTE) academic staff (data provided by the institution)
- ☐ External research income concerning projects on online learning and teaching per Full Time Equivalent (FTE) academic staff (data provided by the institution)
- ☐ Number of visiting scholars per Full Time Equivalent (FTE) academic staff (data provided by the institution)

Teaching staff engaged in research in online teaching & learning
Check all that apply

Please choose **all** that apply:

- ☐ Percentage of Full Time Equivalent (FTE) staff involved in research on online learning and teaching (data provided by the institution)
- ☐ Number of doctorate degrees in online teaching and learning per Full Time Equivalent (FTE) (data provided by the institution)

Research in online teaching & learning - Output
Check all that apply

Please choose **all** that apply:

- ☐ Yearly average n. of publications per Full Time Equivalent (FTE) academic staff (WoS or Scopus publications) (data provided by the institution or review by external panel)

Internationalization
Check all that apply

Please choose **all** that apply:

- ☐ Yearly average number of publications with authors from other countries per Full Time Equivalent (FTE) academic staff (WoS or Scopus publications) (data provided by the institution or review by external panel)
- ☐ Percentage of doctorate degrees that are awarded to international doctorate candidates (data provided by the institution)

Disciplinary research (research groups, research projects, etc.)
Check all that apply

Please choose **all** that apply:

- ☐ Internal budget devoted to disciplinary research per Full Time Equivalent (FTE) academic staff
- ☐ External research income concerning disciplinary projects per Full Time Equivalent (FTE) academic staff
- ☐ Number of doctorate degrees in disciplinary research

Disciplinary research - Output
Check all that apply

Please choose **all** that apply:

- ☐ Yearly average n. of publications per Full Time Equivalent (FTE) academic staff (WoS or Scopus publications) (data provided by the institution or review by external panel)

Comments

Please write your answer here:

Before continuing, please make sure that you selected no more than *6* indicators.

Choose the best indicators for quality of organization

Please choose ***from 0 to 4*** (half) of the following indicators. The indicators you select are those that, in your opinion, are best suited to measure the ***quality of organization***. For each indicator, the data source has been reported in brackets. If you have any comments on the measurable indicators, you can provide them in the text box at the end of this page.

Credit transfer system aligned with national (and /or European) systems and operates bidirectionally

Check all that apply

Please choose **all** that apply:

- ☐ Credit transfer system adopted by the institution (data provided by the institution)

Bureaucratic policies able to cater for the needs of e-learning courses

Check all that apply

Please choose **all** that apply:

- ☐ Operations performable online (checklist: subscription, following lectures, examination, vote registration, ...) (data provided by the institution)

Existence of a complaints and appeals system for learners

Check all that apply

Please choose **all** that apply:

- ☐ Percentage of student complaints or appeals solved/closed (data provided by the institution)
- ☐ Average time (days) for processing complaints/appeals (data provided by the institution)

Bureaucratic support services (providing assistance for admission, financial issues, registration, enrollment,etc.)

Check all that apply

Please choose **all** that apply:

- ☐ Number of full-time equivalents (FTEs) employed for non-instructional, non-technical support services (providing assistance for admission, financial issues, registration, enrollment, etc.) weighted by student satisfaction for the service (data provided by the institution + student survey)

Structures such as libraries, labs, etc.
Check all that apply

Please choose **all** that apply:

- ☐ Student satisfaction for room, laboratory and library facilities (through student survey)

Ability of managing time and avoiding workload
Check all that apply

Please choose **all** that apply:

- ☐ Student paperwork / online forms / front office time burden per week (through student survey)
- ☐ Academic staff paperwork / online forms / front office / commissions time burden per week (through teacher survey)

Student satisfaction
Check all that apply

Please choose **all** that apply:

- ☐ Student satisfaction for organization (through student survey)

Comments

Please write your answer here:

Before continuing, please make sure that you selected no more than *4* indicators.

Choose the best indicators for the sustainability of the institution
Please choose ***from 0 to 3*** (half) of the following indicators. The indicators you select are those that, in your opinion, are best suited to measure the ***sustainability of the institution***. For each indicator, the data source has been reported in brackets. If you have any comments on the measurable indicators, you can provide them in the text box at the end of this page.

Institutional Strategic Plan (ISP) for online education
Check all that apply

Please choose **all** that apply:

- ☐ Availability of an Institutional Strategic Plan for Online Learning (online vision statement, online mission statement, online learning goals and action steps, ...) (data provided by the institution)

Overall coherence of program design and provision (interconnections among courses, flexibility of the design, clarity of program design, ...)
Check all that apply

Please choose **all** that apply:

- ☐ Percentage of curriculum changes resulting from an assessment of student learning (either formal or informal) within a fiscal year [a measure on increased flexibility within the curriculum development process to better respond to a rapidly changing world] (data provided by the institution)

Resources (including financial ones) specifically devoted to the online program
Check all that apply

Please choose **all** that apply:

- ☐ Percentage of total institutional expenditure dedicated to online programmes (data provided by the institution)

Sustainability of the portfolio of programmes
Check all that apply

Please choose **all** that apply:

- ☐ Ratio new students added / lost students per program (in the past year) (data provided by the institution)
- ☐ Ratio new students added in the past year / students still in the program (data provided by the institution)

Clear policy regarding OERs and MOOCs
Check all that apply

Please choose **all** that apply:

- ☐ Percentage of Open Educational Resources used on the total of learning materials (data provided by the institution)
- ☐ Percentage of Massive Open Online Courses on the total number of courses offered (data provided by the institution)

Comments

Please write your answer here:

Before continuing, please make sure that you selected no more than *3* indicators.

Choose the best indicators for the quality of the technological infrastructure

Please choose ***from 0 to 3*** (half) of the following indicators. The indicators you select are those that, in your opinion, are best suited to measure the ***quality of the technological infrastructure***. For each indicator, the data source has been reported in brackets. If you have any comments on the measurable indicators, you can provide them in the text box at the end of this page.

Quality of the overall technological infrastructure

Check all that apply

Please choose **all** that apply:

- ☐ Student satisfaction with the overall learning platform (through student survey)

Robustness of the learning platform

Check all that apply

Please choose **all** that apply:

- ☐ Composite measure of server error rates, average response times, peak response times, and uptime (through technical institutional survey)

Flexibility of the learning platform

Check all that apply

Please choose **all** that apply:

- ☐ Checklist of functionalities supported by the platform (through technical institutional survey)

Adequacy of data security mechanisms

Check all that apply

Please choose **all** that apply:

- ☐ Measure of data security self-assessment (through technical institutional survey).

Compliance of interfaces and contents with usability and accessibility standards

Check all that apply

Please choose **all** that apply:

- ☐ Measure of compliance with the accessibility guidelines WCAG 2.0 (through technical institutional survey)

Existence of a plan for system maintenance and contingency management
Check all that apply

Please choose **all** that apply:

- ☐ Availability of a plan for system maintenance and contingency management (data provided by the institution)

Interoperability of the learning platform
Check all that apply

Please choose **all** that apply:

- ☐ Measure of interoperability (Interoperability with external open sites (e.g., social media, DropBox, Google Drive), interoperability between LMSs (Learning Management Systems), information and teaching/learning materials exchange (LTI, SCORM, ...), Single sign-on (SSO) access control, etc. (data provided by the institution)

Comments

Please write your answer here:

Before continuing, please make sure that you selected no more than *3* indicators.

Any other comments
Please use the section below to add any additional comments or suggestions.

Please write your answer here:

Submit your survey.
Thank you for completing this survey.

Annex 2 – Toolbox testing compilation guide

Dear participant,

Thank you for taking part to the "CODUR Toolbox" testing phase of the CODUR Erasmus+ Project.

The goal of the CODUR project is to provide a tool for the quality assessment of Online Higher Education Institutions. To this end, we devised – with the help of international experts in the field – a list of observable indicators that can be used for assessing and ranking online institutions.

The aim of this phase of the project is testing the *feasibility* of the Toolbox we are proposing. We have devised a list of promising indicators, but before finalizing it, we need to know how much effort gathering these data would require on the part of the institutions.

The goal of the CODUR project is providing a ranking tool that is usable and used: therefore, it is important to make sure that it would be easy to implement.

The "CODUR Toolbox" test spreadsheet

Together with this document, you should have received a spreadsheet containing the list of indicators we devised for assessing the quality of Online Higher Education Institutions. The spreadsheet itself is organized in 8 worksheets (i.e. "tabs") which allow you to easily switch from one criterion to another. Each tab corresponds to one of the eight evaluation criteria proposed by CODUR:

- Tab: "Teaching & Learning";
- Tab: "Student Support";
- Tab: "Teacher Support"
- Tab: "Reputation-Impact";
- Tab: "Research";
- Tab: "Organization";
- Tab: "Sustainability";
- Tab: "Technological Infrastructure".

Each tab is then divided into 10 columns, namely:

- **Criterion**: the CODUR project identified eight criteria on which online institutions should be evaluated. All the indicators we propose refer to a specific criterion, which is reported in this column.
- **Observable indicator**: this is the indicator we propose, for which we ask you to assess the ease of use.

- **Weight**: indicators are weighted differently within each criterion, with some indicators being more important that others. The weight to apply to each indicator was estimated in the previous phases of the project.
- **Data source**: while all data will eventually be provided by the institution, only some indicators will be directly examined by the institution (e.g. the number of student internships). Other indicators will require the institution to administer surveys to students and/or teachers, in order to examine their experience of the institution.
- **Data type**: This column indicates whether an indicator is a raw number, a percentage, the sum score of a survey scale ("Likert scale") or a composite measure of several sub-indicators ("composite scale").
- **How long would it take to collect this data?**: in this column, we ask you to report a rough estimate of the time you would need to collect this data (e.g. half an hour, two days, a week, a semester, …).
- **Does this data need to be collected at a certain time of the year?**: Some data could be more easily collected in a specific time of the year. For example, students surveys may be more easily administered during the months with scheduled lessons, while fiscal data may be more easily collected at the end of the fiscal year. Please report in this column what would be the best time of the year (if any) to measure this indicator in your institution.
- **Has this data already been collected / is this data already available?**: It is possible that you will have some data already at hand, because it was collected for other purposes. If so, please write "yes" in this column.
- **Can this data be made public?**: some of the data we require may be confidential. Please, report in this column whether or not a specific indicator should be considered public, and therefore shareable on our part.
- **Data example**: please, fill this column with an example of what the data from your institution would look like for this indicator. If the data is already available, or doesn't require a lot of effort on your part to collect, report it. Otherwise, if the data type is a raw number or percentage, report a rough estimate based on your experience. The purpose of this column is making sure that the indicator and data type are understandable.

Content of cells from Colum 1 to Column 5 is blocked: this means that their values cannot be edited, but only read. Conversely, content of cells from Colum 6 to Column 10 must be input by you according to the instructions above.

Thank you for your participation!

The CODUR project partners