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

codur

Creating an Online Dimension for University Rankings

PROJECT DELIVERABLE: IO1.A3

Develop test and refine representative performance online quality education indicators based on common criteria

Francesca Pozzi (CNR-ITD), Flavio Manganello (CNR-ITD), Marcello Passarelli (CNR-ITD), Donatella Persico (CNR-ITD)

Date: 19th December 2017
Version 1.1: internally reviewed, added Appendix 4.

Universitat Oberta de Catalunya

Consiglio Nazionale delle Ricerche

Table of contents

1 Introduction

The higher education arena is becoming an ever more competitive market, with universities under constant pressure to secure increasing numbers of students and research funding. It is in this context that university ranking systems have become powerful tools. Systems such as the Times Higher Education World University Rankings[1] and the Academic Ranking of World Universities[2] enable universities, potential students, policy makers and funders to measure and compare universities at a global level (Brasher, Holmes, & Whitelock, 2017).

However, university ranking systems have been criticised (Amsler & Bolsmann, 2012; Lynch, 2015). Furthermore and even more importantly for the CODUR project, although online universities are known to play a crucial role in European Higher Education (McAleese et al., 2014), all of the existing ranking systems presently ignore the specific characteristics of online universities (Brasher et al., 2017). The project acknowledges this gap and puts forwards a number of actions to pave the way to its filling in. 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, such as for example U-Multirank (Van Vught & Ziegele, 2011).

To achieve this goal, during the first year the main tasks of the project have to do with the preliminary design and definition of a set of criteria and indicators. In particular, according to the proposal, the project envisages 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.

In this deliverable, we describe the main activities carried out under IO-A3 and the main outputs.

The document is structured as follows: in Section 2, the overall methodology adopted in Task IO1-A3 is described. In Section 3, the preliminary proposal of the CODUR criteria and indicators is illustrated, along with the process that led the consortium to its definition. In Section 4, we describe the two actions used so far to collect external feedback and reactions to the CODUR criteria and indicators, i.e. the Delphi Study and the EMEM workshop. Both actions are described and their respective outputs are reported. In Section 5, we discuss the main lessons learnt and provide recommendations for future work. Section 6 reports the main conclusion of this work.

1

2

2 A participatory methodology for the definition of the CODUR criteria and indicators

As already mentioned, the main goal of the project is to propose a set of criteria and indicators for the evaluation of online Higher Education (HE) institutions, to be possibly integrated with existing ranking systems.

In line with what is stated in the CODUR proposal, the project takes a participatory approach to the definition of the criteria and indicators, i.e. the design phase is not something that happens within the project boundaries, but, on the contrary, it involves several stakeholders and informants and the broader HE community. This is done in the assumption that taking into consideration the points of view of all the relevant informant 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.

This is the reason why the project starts from the analysis of what already exists in terms of systems and tools for *traditional* HE, puts forward a preliminary project proposal which focuses on *online* HE and then – thanks to different means and actions – involves a number of people and bodies external to the project for their iterative revision, testing and valorisation.

In particular, Figure 1 illustrates the overall methodology of Task IO1-A3.

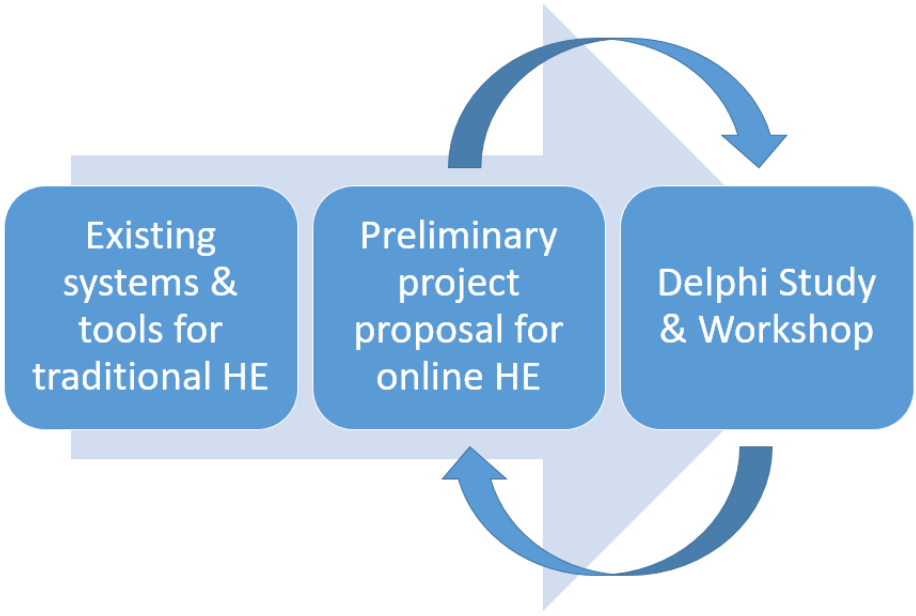


Fig. 1 – Overall methodology of Task IO-A3

As it is illustrated in Figure 1, the project started from the analysis of the existing quality assurance tools and systems (carried out under Task IO1-A2), along with the analysis of the current provision of online higher education institutions worldwide (Task IO1-A1). Building on these two inputs, the project partners elaborated in a collaborative way a preliminary set of criteria and indicators.

This preliminary set was then refined, enriched, tested and valorised by means of:

- a Delphi Study, involving a number of experts worldwide

- a face-to-face workshop held in Italy, involving stakeholders from the national HE community.

These two activities have been used to collect feedback, comments and new ideas about the preliminary set of criteria and indicators, with the aim to inform their upcoming revision.

To be noted that the Delphi Study was not originally mentioned in the CODUR proposal, but it was the method of choice agreed upon by the project partners, as the use of this technique is a valuable means to involve at a distance a worldwide community of experts and lead them to reach a consensus on the proposed criteria and indicators.

3 Preliminary proposal of criteria and indicators

As already mentioned, building on the inputs gained from IO-A1 and IO-A2, the consortium elaborated a preliminary set of criteria and indicators for online HE institutions.
This was done through several interactions among the partners during Year 1 of the project and in the context of the transnational project meeting held in Genoa in June 2017, during which a specific session was devoted to the collaborative systematization of the criteria and indicators.
The Genoa meeting provided the project partners with a fundamental opportunity to discuss in presence about the state of their activities with regard to the following project tasks:

- presentation of the state of the art of Online Education worldwide (IO1-A1),
- discussion about the existing ranking tools and quality assurance systems and their main indicators (IO1-A2).

In particular, presentation and discussion of both IO1-A1 and IO1 -A2 served to set up a common ground for the systematization of a shared and agreed preliminary set of the CODUR criteria and indicators.

In the following section, the method used in Genoa to agree on the preliminary set of criteria and indicators is described, and in section 3.2 the result of such activity is reported.

3.1 Method

In order to define and then refine a first shared version of the CODUR criteria and indicators among the project partners, during the Genoa meeting a specific session was devoted to a collaborative activity.

In order to reach that point, the technique used during the activity was the Metaplan[3]. The technique is quite famous and it is often used within collaborative working and decision taking contexts, because it takes a participatory approach and facilitates discussion and opinion exchange, especially during design activities. These characteristics seemed to fit well with the project needs.

Based on the preliminary discussion and the information shared among partners during the meeting, the Metaplan activity was organized in the following phases:

- Preparation phase: before the activity started, one poster was hung on the wall entitled “Criteria & Indicators”. Besides, each participant was equipped with a set of coloured post-its.
- 1st phase: individual work: during this phase each participant was asked to write down on his/her post-its all the criteria and indicators s/he deemed to be important and then hung them on the wall.
- 2nd phase: collaborative discussion: during this phase the group was asked to consider all the post-its and cluster them, eliminate duplicates, agree on the level of granularity of the clusters, thus distinguishing criteria from indicators, etc. (see Fig. 2).

Overall, the activity was quite productive, as several aspects emerged during the discussion and many important decisions were taken.

3

First important aspect that emerged during the activity and that was definably fixed, has to do with the focus addressed by the project: even if in the proposal it is clearly stated that the expected criteria and indicators would address online Higher Education institutions, the need emerged quite early to fix this more precisely, as sometimes the discussion tended to focus on lower levels of granularity, addressing for example the evaluation of courses or programmes (rather than institutions). So, the consortium concluded these two latter factors are out of the scope of the project, which instead addresses the evaluation of *institutions*.

Another important aspect that was fixed during the Metaplan, has to do with the terminology. It was necessary to clearly define terms, such as (online) dimension, criterion, indicator, category, cluster, that were sometimes used in a confusing way. Thus, the project consortium agreed on how to order these terms, as it is illustrated in Fig. 2.

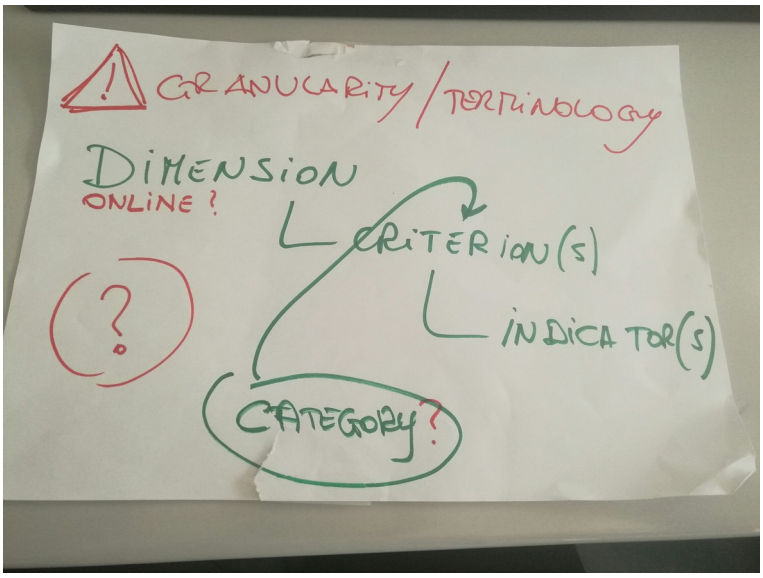
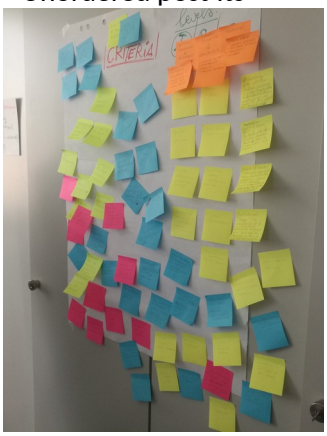


Fig. 2 – Consensus reached on the terminology

In particular, we agreed to use the term “dimension” to indicate the higher level; in order to evaluate the online dimension of an HE institution, we will define a set of “criteria” (possibly aggregated into broader categories); each criterion will be then measured by means of lower level “ indicators”.

In Figure 3, we can observe how the structure of the post-its evolved: from an unordered set of post-its hung by each one during the individual working phase (a), after the discussion the post-its could be easily arranged and took the form of three different threads (clusters) (b).

a) Unordered post-its



b) Ordered post-its (3 clusters)

Fig. 3 – Defining the preliminary set of criteria and indicators

Quite interestingly, during the discussion it was noted that the three clusters somehow reflected the three categories put forward by Ossiannilsson, William, Camilleri, & Brown (2015), namely: Services, Products and Management.

In the following section, the complete set of criteria and indicators is reported.

3.2 Output

As already mentioned, in line with the proposal put forward by Ossiannilsson et al. (2015) and with what it is recommended in IO1-A2 (Brasher et al., 2017), our preliminary set of criteria and indicators is structured into three main categories: ***Services***, ***Products***, ***Management*** (see Table 1). Such categorization emerged quite naturally during the collaborative activity of the Metaplan.

Table 1 – Preliminary set of CODUR criteria and indicators

SERVICES
Criterion: Quality of student support [DEF: it refers to the ability of the online higher education institution to provide support to learners in many different areas, for example learning, orientation, socializing with peers, organisational issues, use of technology, etc.] *Examples of Indicators:* ▪ Orientation services to help learners taking decisions about their learning path ▪ Mentoring, tutoring activities to support student learning ▪ Support to student community building ▪ Support to alumni community building ▪ FAQ systems, Helpdesk, or sim. for learners ▪ Newsletters or sim. to allow communication from the institution to the learner ▪ Regularity of information update ▪ Technology support ▪ …
Criterion: Quality of teacher support [DEF: it refers to the ability of the online higher education institution to provide support to teachers and lectures in terms of training provision, organisational issues, use of technology, etc.] *Examples of Indicators:* ▪ Opportunities for teaching staff training in online education ▪ Support to teachers and lecturers (for example provided via a specific department devoted to this) ▪ Support to teacher community building ▪ FAQ systems, Helpdesk or sim. for teachers ▪ Newsletters or sim. to allow communication from the institution to the teaching staff ▪ Regularity of info update ▪ Technology support ▪ …

Criterion: Quality of technological infrastructure
[DEF: it refers to the ability of the online higher education institution to offer a sound technological platform, in terms of usability, accessibility, flexibility, types of features offered, etc.]

Examples of Indicators:

- Compliance of interfaces with usability and accessibility standards
- Consistency of the platform
- Capacity and concurrence of the platform (i.e. number of people that the infrastructure can host, even at the same time)
- Flexibility and scalability of the platform in view of future changes /new emerging needs
- Capability of the platform to satisfy different user needs (personalization capabilities, etc.)
- Capability to support interaction and teamwork
- Capability to support evaluation
- Adequacy of the technical support (provided by department/ ad hoc team/or external company in charge of the technology)
- Adequacy of data security mechanisms
- Existence of a plan for system maintenance and contingency management
- …

PRODUCTS

Criterion: Quality of learning experience
[DEF: it refers to the ability of the online higher education institution to offer effective learning experiences, in terms of design, delivery, methods, learning materials, assessment means, etc.]

Examples of Indicators:

- Quality of course / learning design
  - Existence of syllabus
  - Availability of information before the course begins
  - Definition of clear objectives
  - Clear definition of contents to be addressed
  - Clear definition of the assessment strategies
  - Suitability of methods to learning objectives
  - Possibility to personalize the learning path
  - Consistency of the estimated time/workload
  - Consistency of the expected outcomes
  - Availability of data to check the student progress
  - ….
- Quality of pedagogy /methodology
  - Adoption of innovative approaches to online pedagogy
  - Level of interactivity of learning activities
  - Adoption of collaborative learning approaches
  - Activity variety (individual work, team work, problem based work, simulations, reviews, case studies, etc.)

- Support provision to self-regulation (including self- monitoring, self-assessment, etc.)
- Support provision to personalization
- …

- Quality of assessment
  - Assignments and assessment tests are aligned with the learning objectives
  - Diversity of assessment approaches (quantitative, qualitative, etc.)
  - Diversity of assessment tools (human-based, technology- based, etc.)
  - Availability of formative assessment
  - Availability of anti-cheating mechanisms/tools
  - …
- Quality of learning materials
  - Rigourness of materials
  - Suitability of materials to contents and learning objectives
  - Regularity of updating
  - Variety of learning materials (formats, types, etc.)
  - Level of interactivity of materials
  - Clear policy about the intellectual property of the learning materials
  - Materials are compliant with accessibility standards
  - Contents and materials take into account diversities (cultural, gender, etc.)
  - ….

Criterion: Quality of research
[DEF: it refers to the ability of the online higher education institution to carry out research initiatives and innovation projects, for both disciplinary research and research on online teaching and learning]

Examples of Indicators:

- Research output (in terms of publications, visiting scholars, …..)
- Research in online teaching & learning (research groups, research projects, etc.)
- Teaching staff engaged in research in online education
- Internal centres devoted to research in online education
- Research in other disciplines
- Use of research for improvement and innovation (research based projects…)
- ….

MANAGEMENT

Criterion: Organization
[DEF: it includes aspects such as availability of structures providing services of various kind, lightness and efficiency of bureaucracy, etc.]

Examples of Indicators:

- Decentralized structures on the territory (i.e. having physical structures accessible to students enrolled in the programmes)
- Structures such as libraries, labs, etc.
- Non instructional support services (providing assistance for admission, financial issues, registration, enrollment,etc.)
- Examination policies able to cater with for the needs of e-learning courses

- Credit transfer system aligned with national (and /or European) systems and operates bidirectionally
- Existence of a complaints and appeals system for learners
- …

Criterion: Quality of teaching
[DEF: it refers to the ability of the online higher education institution to recruit experienced teachers trained in delivering online teaching, provide them with standards for teaching, guarantee regular quality control procedures, etc.]

Examples of Indicators:
- Profile of the teaching staff
- general (knowledge field based) qualification
- knowledge field experience
- online qualification
- (online) teaching experience
- Teacher assessment and quality control
- Involvement of academics in teaching
- Standards for regulating teacher-student interactions (teachers' response time, modalities, timeliness of feedback provision for tests, etc.)
- …

Criterion: Sustainability of the institution
[DEF: it includes aspects such as the size of the institution, resources, availability of standardised procedures and strategic plans, etc.]

Examples of Indicators:
- Institutional strategic plan for online education
- Size of the institution
    - staff size (number of teachers, lecturers, researchers, admin staff, …)
    - number of students enrolled
    - number of programs
    - number of courses
    - number of departments
    - number of disciplines covered
    - …
- Quality of overall program design and provision (interconnections among courses, flexibility of the design, clarity of program design, …)
- Quality management procedures for program /course /material design and development
- Sustainability of the portfolio of programmes
- Resources (including financial ones) specifically devoted to the online program
- Clear policy regarding OERs and MOOCs
- …

Criterion: "Reputation" /impact
[DEF: it includes aspects such as impact on the job market, institutional image, communication strategies, etc.]

Examples of Indicators:
- Job opportunities and impact of developed competencies and skills on the job market

- Internship and mobility opportunities
- Social impact (same of other Universities or other?)
- Institutional image (e.g. position in other ranking systems, or existence of institutional merchandise)
- Communication strategies (e.g. use of academic social networks)
- Organization of events, conferences, etc.
- …

Of course, the list is far from being complete and the proposed criteria and indicators are not always clear-cut, since this was only a first draft that needed to be further nurtured, enriched and improved. However, at the end of the Genoa meeting, partners agreed this was a good starting point that could be used as a valuable input for the following activities of the project.

In the following section, we describe how this input was further used and exploited by the project.

4 Refining and valorising the CODUR criteria and indicators

In this section, we present the two main actions carried out to refine, enrich and valorise the preliminary criteria and indicators illustrated in the previous section.

In particular, Section 4.1 describes the Delphi Study, while Section 4.2 describes the face-to-face Italian workshop.

4.1 The Delphi Study

The Delphi method is a research technique based on consultation with a panel of experts. Delphi Study as a research method originated in the 1950s by the RAND Corporation (Research ANd Development) and it was first described in research in Dalkey & Helmer (1963). In general, the Delphi studies are intended to predict trends, problems and possible developments of a given (research) sector (Rowe & Wright, 1999). With several variants, Delphi is used to construct possible short and medium range scenarios on a wide variety of themes. On the other hand, Delphi can also be used without this predictive function as a simple analysis tool based on experts' judgment.

As discussed by Plesch et al. (2012), Delphi studies have some key characteristics: questionnaires are proposed in an iterative way, aiming at gathering information about the experts opinions on future trends on a field of interest; usually all the experts participating in the study remain anonymous; the results of the previous rounds are used to prepare questionnaires for the following rounds.

The fundamental characteristic of Delphi is that the various experts included in the panel are ignorant of the identity of the other participants during the process. Each expert works independently and the interaction is mediated by a conductor (or facilitator) who receives the individual contributions, synthesizes them and redistributes them to the experts anonymously in order to obtain further information through subsequent Delphi rounds (Linstone & Turoff, 1975). Usually, the experts answer questionnaires in two or more rounds. As a technique where there is no direct interaction between group members (i.e. the group is *nominal*), there may be many participants. One point of weakness is the wide range of opinions to be elaborated and synthesized by the conductor. For this reason, the role of the conductor is central and crucial: she/he must be able to interpret correctly and comprehensively the individual contributions of the participants. A second critical aspect is that, when rounds are carried out with the same set of experts, the number of respondents often decreases drastically. Again, the role of the conductor is critical in identifying experts that are committed to the task in a long term perspective.

As already stated before, there are several forms of Delphi although we can recognize two main structures and two main purposes: while one tends towards the agreement of the participants, which however is not taken for granted and is not forced, the other does not pursue consensus, even though this can be appreciated. The conclusion of a Delphi can consist of one or more theses of a majority of participants with variants and differentiations of a minority.

General description of Delphi

The preliminary phase of a Delphi consists in recruiting the experts for the panel, sending them general information on how the method works and introducing the activities. Then, the panel's members receive a first questionnaire to be answered individually within the scheduled agenda. The facilitator collects the various responses and organizes them in a synthesis that will be sent back to the experts along with a second questionnaire. Depending

on the purpose of Delphi, the second questionnaire may be similar to the first or more specific (funnel structure). Participants can read the synthesis, possibly excerpts from others' answers (always reported in anonymous form), get an idea of how the group is arguing, etc. but they cannot trace the identity of the expert who has expressed a particular position. This peculiar way of managing the group during a Delphi is designed to handle, for example, different well-known conflicting positions, which could create negative dynamics within the group and limit the spontaneity of the interactions. At the end of the last round (usually two rounds are envisaged), the facilitator sends the summary to all the participants.

4.1.1 Method

Within the CODUR project, a Delphi Study was designed as a structured procedure relying on a nominal group (i.e. the panel of experts). It was used as an analysis tool based on judgment of experts, rather than as an interactive forecasting method. The aim of the CODUR Delphi Study is to involve at a distance a worldwide wide community of experts and allow them to reach a consensus on the proposed criteria and indicators. In particular, the objective is to interrogate a panel of experts for gathering their views on indicators for evaluating the quality of online higher education institutions.

An overview of the CODUR Delphi Study research design is presented in the diagram below (fig. 4) and better described as following.

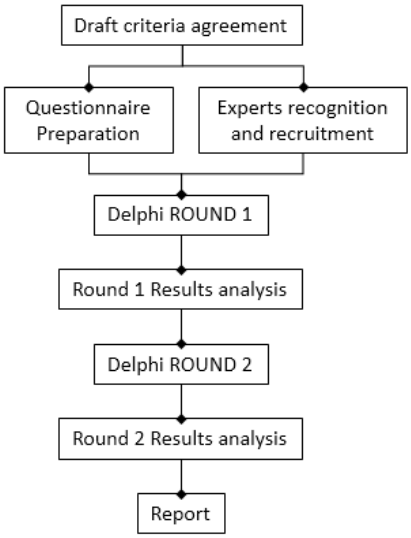


Fig. 4 – CODUR Delphi Study Research Design

The preliminary proposal of criteria and indicators - elaborated as described in section 3 of the present document - was used as the starting point for the Delphi Study Research design (*draft criteria agreement*) and a *questionnaire* was prepared. At the same time, a set of criteria for selecting the expert panelists was agreed, in terms of roles and number of stakeholders to be targeted, and a list of experts to be contacted was elaborated and shared among the project partners (*experts recognition and recruitment*).

The experts are expected to react to the questionnaire in two rounds, each of which goes through a process of analysis (made by the project research team) and returns to the experts in the form of an anonymized summary.

Finally, the process is stopped and the results of the Delphi Study are presented in a *report*.

As already mentioned, the expected result of the CODUR Delphi Study is the achievement of consensus on the proposed criteria and indicators, but the qualitative process of analysis that the research team carries out is also expected to valorize significant levels of possible emergent disagreement (i.e. *variants and differentiations of a minority*).

The detailed workflow of the activities carried out by the partners so far (from the Draft criteria agreement to Round 1) is presented in the diagram below (Fig. 5).

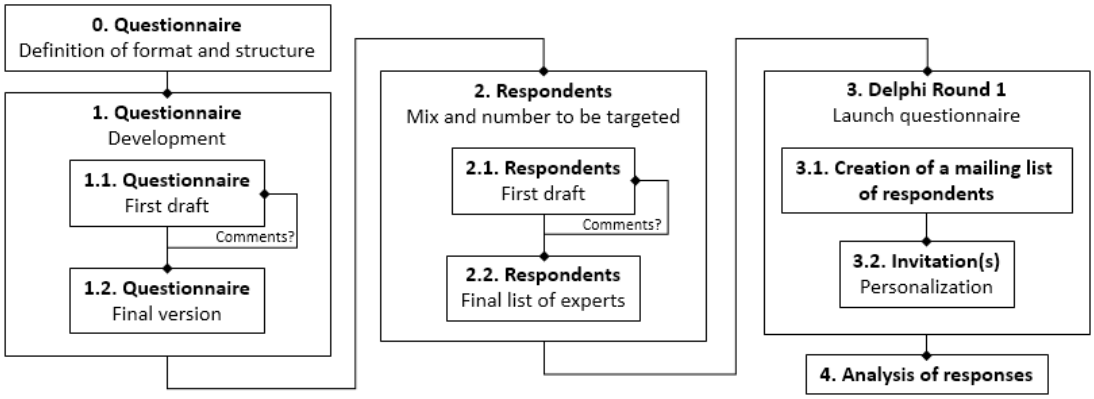


Fig. 5 – CODUR Delphi Study Plan (to Round 1)

The "CODUR - Towards the creation of an online dimension for University Rankings" *questionnaire* was defined by the project partners both in the format and in the structure. The process toward the *development* of the questionnaire was managed in a collaborative and iterative way. This process started by sharing online a *first draft* of the questionnaire, and then by means of a collaborative work of comments and edits a *final version* was released.

The final version of the questionnaire is aligned with what is described under I01-A3 of the project proposal. In particular, the core section of the questionnaire aims to reach the four main objectives declared in the task, i.e.:

- to identify the criteria relevant for on-line educational institutions,
- to determine a non-ambiguous, agreed-upon definition for each criterion,
- to operationalize each criterion, by identifying multiple observable indicators,
- to assign a relative weight to each criterion.

The criteria for selecting and recruiting the respondents were defined. The project partners agreed to consider as expert panelists a selected group of interested stakeholders (including students, teachers and researchers). A *final list of experts* and informants was compiled and shared. For the purpose of the CODUR Delphi Study, students considered experts are those able to provide the project with meaningful information about the student viewpoint (i.e. representatives of Student Associations, QA student expert pool, ...). Results from student participants are discussed separately, in Appendix 4.

During Round 1, the questionnaire presented the experts with a number of indicators for university raking systems, clustered in nine criteria. The experts were asked to evaluate the importance of these criteria and indicators based on their professional experience. In line with the proposal put forward by the partners (see Section 3), the criteria are: Quality of student support, Quality of teacher support, Quality of technological infrastructure Quality of learning experience, Quality of research, Organization, Quality of teaching, Sustainability of the institution, Reputation/ "impact". For each of the nine criteria, participants were

presented with a list of indicators and they 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). In the questionnaire, the nine criteria were presented in a random order.

A copy of the questionnaire used can be found in Appendix 1.

In order to run the Delphi Study via Internet, LimeSurvey - a free and open source online survey application – was used to create and *launch the questionnaire*. LimeSurvey enables the CODUR research team to collect responses, create statistics, and export the resulting data for the *analysis of the responses*.

The following main actions are planned at the end of the Delphi Round 1: to analyze the data about all the responses collected; to summarize the results into an anonymized summary; to prepare detailed questions for the *Delphi Round 2*.

4.1.2 Participants

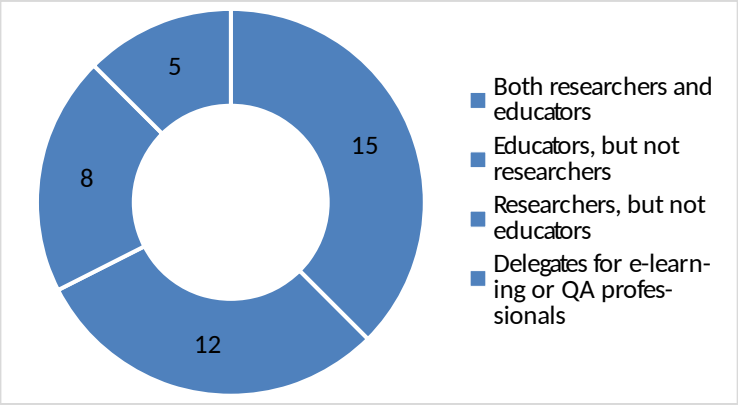
The panel of experts consists of a group of interested stakeholders selected by the project partners. The list of panelists include students, teachers and researchers – at an international level – who are considered informed and competent on the topic.

A number of 140 experts was identified and a selected sub-set of them was contacted to take part to the early stage of the Delphi Round 1. The process of experts recruiting is managed in two steps: a first email is sent to formally invite the expert to participate to the Delphi Study; a second email is sent if the expert agree to participate, in order to provide her/him with all the information and the link to the questionnaire.

At the current stage, 40 participants took part in Round 1 of the CODUR Delphi Study (17 females, 19 males, 4 undisclosed; age 53.95 ± 9.47, range 35-72)[4].

Twenty-one participants reported Italy as their country, while the rest of the sample reported Australia (6), Spain (6), the UK (3), or Israel, Canada, or Bulgaria (1 each). One participant did not disclose his or her home country.

Fifteen participants reported being both researchers and educators; 12 are educators, but not researchers; 8 are researchers, but not educators. The remaining five reported being either delegates for e-learning or QA professionals. (Fig. 6). To be noted that Round 1 was started during summer time; this limited the participation of students in this phase. Additional data from students was collected at a later date and results from their participation are presented and discussed separately, in Appendix 4.



4 The list of participants cannot be made public at this stage of the process, as Round 2 is ongoing. The list will be made available at the end of the Delphi Study.

Fig. 6 – CODUR Delphi Round 1 - Respondents background

4.1.3 Output

In the following section, we describe the results of the data collected with the first round of the Delphi.

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. As already mentioned, rating was performed on a 0-4 scale (from 0 = not at all important to 4 = extremely important). Participants were allowed to choose "unsure / unclear" if they did not feel confident in rating an indicator (which has been coded as a missing value). The nine lists of indicators were randomized in order across participants.

The quartile distribution for these ratings are reported in Table 2.

Table 2 – Indicators' importance

Criterion	Indicator	1st quartile	Median	3rd quartile
Quality of student support	Orientation services to help learners taking decisions about their learning path	3	4	4
	Mentoring, tutoring activities to support student learning	4	4	4
	Support to student community building	2.5	3	3
	Support to alumni community building	2	2	3
	FAQ systems, Helpdesk, or sim. for learners	2.5	3	4
	Newsletters or sim. to allow communication from the institution to the learner	2	3	3
	Regularity of information update	3	3	4
	Technology support	3	4	4
Quality of teacher support	Opportunities for teaching staff to be trained in online education	3	4	4
	Support to teachers and lecturers (for example provided via a specific department devoted to this)	3	4	4
	Support to teacher community building	2	3	4
	FAQ systems, Helpdesk or sim. for teachers	3	3	4
	Newsletters or sim. to allow communication from the institution to the teaching staff	2	3	3
	Regularity of info update	2	3	4
	Technology support	3	4	4
Quality of technology	Compliance of interfaces with usability and	3	4	4

	accessibility standards			
	Consistency/ robustness of the platform	3.5	4	4
	Capacity and concurrence of the platform (i.e. number of people that the infrastructure can host, even at the same time)	3	4	4
	Flexibility and scalability of the platform in view of future changes /new emerging needs	3	4	4
	Capability of the platform to satisfy different user needs (personalization capabilities, etc.)	3	4	4
infrastructure	Capability to support interaction and teamwork	3	4	4
	Capability to support evaluation	3	4	4
	Adequacy of the technical support (provided by department/ ad hoc team/or external company in charge of the technology)	3	4	4
	Adequacy of data security mechanisms	3	4	4
	Existence of a plan for system maintenance and contingency management	3	4	4
	Quality of course / learning design	4	4	4
Quality of the learning experience	Quality of pedagogy /methodology	4	4	4
	Quality of assessment	4	4	4
	Quality of learning materials	3.5	4	4
Quality of research	Research output (in terms of publications, visiting scholars, ….)	3	3	4
	Research in online teaching & learning (research groups, research projects, etc.)	3	3.5	4
	Teaching staff engaged in research in online education	3	3	4
	Internal centres devoted to research in online education	3	3	4
	Research in other disciplines	2	3	3
	Use of research for improvement and innovation (research-based projects…)	3	4	4
Quality of teaching	Profile of the teaching staff	3	4	4
	Teacher assessment and quality control	3	3	4
	Involvement of academics in teaching	3	3	4
	Standards for regulating teacher-student interactions (teachers' response time,	3	4	4

	modalities, timeliness of feedback provision for tests, etc.)			
Quality of organization	Decentralized structures on the territory	1	2	3
	Structures such as libraries, labs, etc.	2.5	3	4
	Non instructional support services (providing assistance for admission, financial issues, registration, enrollment,etc.)	2.5	3	4
	Examination policies able to cater with for the needs of e-learning courses	3	4	4
	Credit transfer system aligned with national (and /or European) systems and operates bidirectionally	3	4	4
	Existence of a complaints and appeals system for learners	3	3	4
Sustainability of the institution	Institutional strategic plan for online education	3	4	4
	Size of the institution	2	2	3
	Overall coherence of program design and provision (interconnections among courses, flexibility of the design, clarity of program design, ...)	3	4	4
	Existence of standardized workflows for program /course /material design and development	3	3	4
	Sustainability of the portfolio of programmes	3	3	4
	Resources (including financial ones) specifically devoted to the online program	3	4	4
	Clear policy regarding OERs and MOOCs	2	3	4
Quality of reputation	Job opportunities for graduates	3	3	4
	Internship and mobility opportunities	2.5	3	4
	Social impact	3	3	4
	Institutional image	3	3	3
	Communication strategies	2	3	3
	Organization of events, conferences, etc.	2	2	3

Examining the results, we can see that most indicators are deemed very important for assessing their criterion. The only indicators with a median below 3 - which could still be considered a high threshold - are the support to alumni community building (student support), presence of decentralized structures on the territory (organization), size of the

institution (sustainability of the institution), and organization of events, conferences, etc. (quality of reputation).

Moreover, there is agreement between raters for indicator importance: a modification of Fleiss' kappa for ordinal data (Marasini, Quattro, & Ripamonti, 2014) reports an agreement of 0.55 (percentile-bootstrapped 95% confidence interval of [0.52, 0.58]), which can be considered a moderate agreement.

In the survey, participants were also allowed to propose additional indicators for each criterion, if they deemed it appropriate. As a result of this, several indicators we didn't consider were suggested.

The following list omits indicators we already considered for a different criterion (e.g. quality of teaching has been suggested as an indicator of reputation, but is already its own criterion).

- For Quality of student support, suggested indicators were: encouraging peer learning; possibility of personalisation and self-regulation opportunities; dematerialization of processes and smart information system use (e.g. smartphone app).
- For Quality of teacher support, suggested indicators were: recognition of teachers´ efforts in their career (e.g. tenure, economic advantages); availability of good quality sample courses with increasing levels of teacher involvement; professional development systems to support professional learning on teaching and learning in higher education; updating of support materials.
- For Quality of technological infrastructure, proposed indicators included: interoperability with external open sites (e.g., social media, DropBox, Google Drive); technician training; compliance with international standards for interoperability between LMSs (Learning Management Systems), information and teaching/learning materials exchange (LTI, SCORM, ...), and Single sign-on (SSO) access control; ability for learners and teachers to innovate and use newer add-ons/plugins; the cost of providing the infrastructure versus the cost of providing people to support student interaction with the infrastructure.
- For Quality of the learning experience, new suggestions included: quality of interactions between educators and learners; specific training courses for teachers accessible throughout the academic year; quality of online tutoring/mentoring/assistance; tracking of online interactions; quality of interactive environments and collaborative environments.
- For Quality of research, new propositions were: involvement of students in research activities as co-partners; percentage of staff that is active in research; interdisciplinary impact; effective application of research findings and best practices in teaching.
- For Quality of organization, participants suggested to consider: ability of managing time and avoiding workload; teachers' collaboration with external partners, experts, institutions to deploy innovative and authentic learning settings; teachers' interdisciplinary and technical collaboration in forms of co-teaching and co-working; student satisfaction.
- For Sustainability of the institution, suggestions were: clear indications about sustainability issues when evaluation is concerned; pedagogical training of teacher and training concerning the code of ethics; connections with other institutions at local level; interconnections with global institutions; funding.
- Lastly, for Quality of reputation participants suggested to consider: placement in international and national rankings; relationship with the territory; visibility on

academic social networks; number of students; attrition rate; number of students completing the courses; standard of assessment; representation on national forums; existing traditions of the institution; public profile of staff, students, and alumni; credibility of the institution.

Some indicators were proposed for multiple criteria (e.g. relationship with local communities and institutions has been proposed for both reputation and sustainability). Further considerations should be focused in determining whether or not these indicators should influence both criteria, and if not, which criterion they should be assigned to.

Relative importance of criteria

In order to obtain a first estimate of the weight that should be assigned to each criterion when ranking online institutions, participants were asked to order the nine criteria from the most important to the least important.

The criteria rankings thus obtained were analyzed using Thurstone Case V scaling (Thurstone, 1927). This kind of analysis allowed us to obtain estimates of importance on an arbitrary scale, complete with 95% confidence intervals (see Figure 7).

Additionally, asking participants to rank criteria, instead of judging the importance of each one separately, avoided the possibility that all criteria would be ranked as very important.

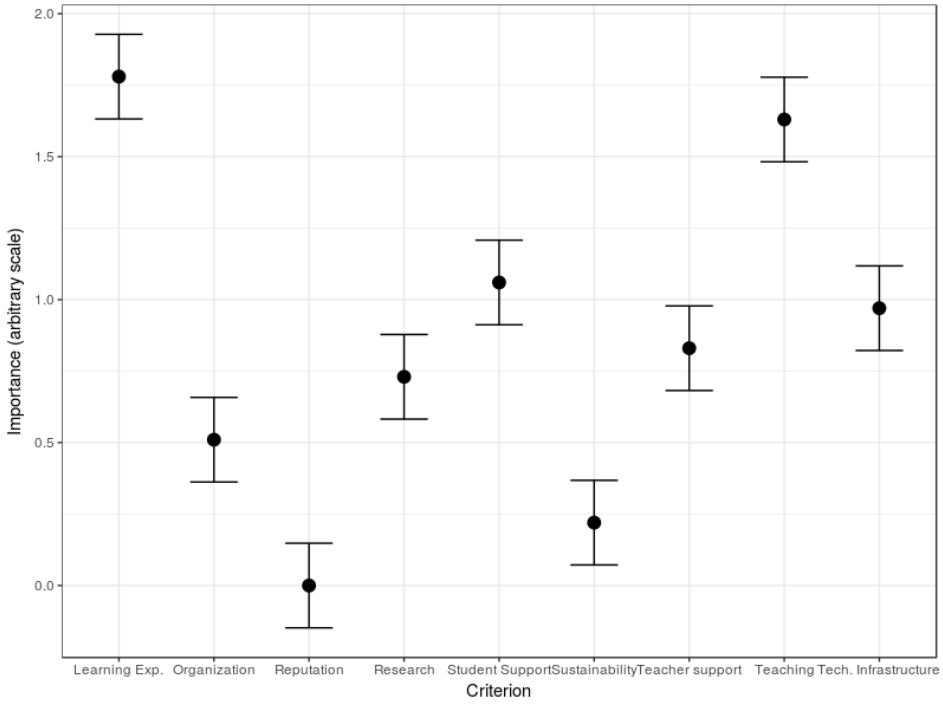


Fig.7 - Relative importance of the evaluation criteria

Looking at the obtained scaling, we can see that Quality of the learning experience and Quality of teaching have generally been considered the two most important criteria for evaluating online institutions.

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.

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

Lastly, it should be noted that the first part of the survey asked participants to judge the importance of indicators for measuring their criterion independently on the perceived importance of the criterion (e.g., even if reputation of the institution is considered of little importance, the indicators proposed could very well be important for measuring Reputation; conversely, even if Quality of teaching was considered extremely important, the proposed indicators could in theory be inappropriate).

However, participants seem to have been at least partly influenced by the importance they assigned to the criterion itself. The criterion weights obtained from the Thurstone Scaling correlate .40 ($p < .001$) with the judgements of importance of their indicators, suggesting the existence of a linear relationship between these two kinds of judgements.

4.2 The EMEM workshop

The EMEM conference[5] is an Italian event organized each year by the SIe-L[6] (the Italian Association for e-Learning). It gathers the most relevant stakeholders and academics working in the field of e-learning, especially as far as Higher Education is concerned. This year the event was held in Bolzano, from August, 30th till September, 1st. In the CODUR proposal, the event had been already identified as a potential arena for the Italian Multiplier Event and for this reason the consortium submitted a proposal for a workshop to be held within the conference context. The proposal was successful and consequently the workshop was held on September, 31st. The title of the workshop was: "Verso il riconoscimento della 'dimensione e-learning' nei sistemi di ranking delle Università" ("Towards the recognition of the e-learning dimension in the University ranking systems").

In the following, the main workshop characteristics are described.

Workshop objectives

The workshop main aim was to promote debate and exchange about the evaluation of the online dimension within the University ranking systems. To reach this aim, the workshop was organized in two phases:

- in the morning, a Round Table was held where six invited experts gave their contributions on the topic with great interest for the participating audience;
- in the afternoon, a working session was organized as continuation of the morning session, where all the workshop participants (experts and audience from the morning session) could actively contribute and provide their opinions and feedback.

Target and recruitment of participants

The workshop expressly targeted all the Higher Education stakeholders (teachers, researchers, students, Ph.Ds, etc.) with an interest in the debate around the quality of online teaching and learning within the University context.

5

6

The workshop was advertised through the main conference website[7] and mailing list, as well as through other channels activated directly by CNR-ITD (news on the ITD website[8], use of institutional social media, use of institutional mailing lists, etc.).

Structure of the workshop

As already mentioned, the workshop was organized into two phases.

In the first phase a Round Table was held, which brought together six Italian experts who were asked to provide their opinion about the workshop topic ("Towards the recognition of the 'e-learning dimension' in the University ranking systems"). Prior to the event, the experts had been provided with the provisional list of the CODUR criteria and indicators (see Section 3) and were asked to comment on them during their speech.

During the second phase, all the workshop participants, including both the experts and the audience of the Round Table, were actively engaged in a decision-making collaborative session, through which we collected further inputs and feedback concerning the CODUR criteria and indicators.

For a more detailed description of the workshop activities, see Section 4.2.2.

4.2.1 Participants

Overall, the workshop was attended by 38 persons. Even if we do not have detailed profiles of the participants, knowing the traditional audience of the EMEM conference, we can assume most of them were teachers (from both University and school) and researchers.

The experts invited to the Round Table are all very well-known policy makers in the Italian e-learning context, as they are all members of the Sie-L (Società italiana di e-learning, *Italian e-Learning Association*) Steering Committee. In particular:

- *Floriana Falcinelli*, from the University of Perugia, is University Deputy for e-Learning.
- *Patrizia Ghislandi*, from the University of Trento, is Director of the University Lab for Educational Innovation.
- *Pier Paolo Limone*, from the University of Foggia, is University Deputy for e-Learning.
- *Tommaso Minerva*, from the University of Modena and Reggio, is member of the University Committee for e-learning.
- *Pier Giuseppe Rossi*, from the University of Macerata, is University Deputy for e-Learning and Director of the University Centre for e-learning (CELFI).
- *Marina Rui*, from the University of Genova, is University Deputy for e-Learning.

4.2.2 Workshop report

The Round Table was held during the morning session and lasted from h. 11.10 a.m. until h. 01:00 p.m.

After an introductory presentation (see Appendix 2), held by Francesca Pozzi (CNR-ITD), aimed to introduce the main topic of the workshop, along with the CODUR project and the proposed list of criteria and indicators, the floor was left to the experts (see names above).

The Table was chaired by Donatella Persico (CNR-ITD), who allowed two rounds of opinions for each expert.

7

8

The discussion turned out to be quite alive, thus demonstrating that the proposed topic is very hot within the Italian University context.

During the Table, we collected a significant feedback related to the CODUR proposal that are described in detail in Section 4.2.3.

Fig. 8 – The Round Table (experts and audience)

As already mentioned, the afternoon session was devoted to a collaborative activity, involving all the workshop participants. To this aim, two groups were formed. The session started at 2:00 p.m. and finished at 3:50 p.m.

At the beginning of the afternoon, a discussion among all the participants emerged very naturally, as a follow up of the morning Round Table: given that during the morning session there had been little time for the audience to intervene and ask questions to the experts, this was done quite spontaneously at the beginning of the afternoon session. This is also a good indicator of the topic to be perceived by the workshop participants as quite relevant.

After this spontaneous discussion, the collaborative activity was launched as planned. Each group was equipped with a set of paper-forms:

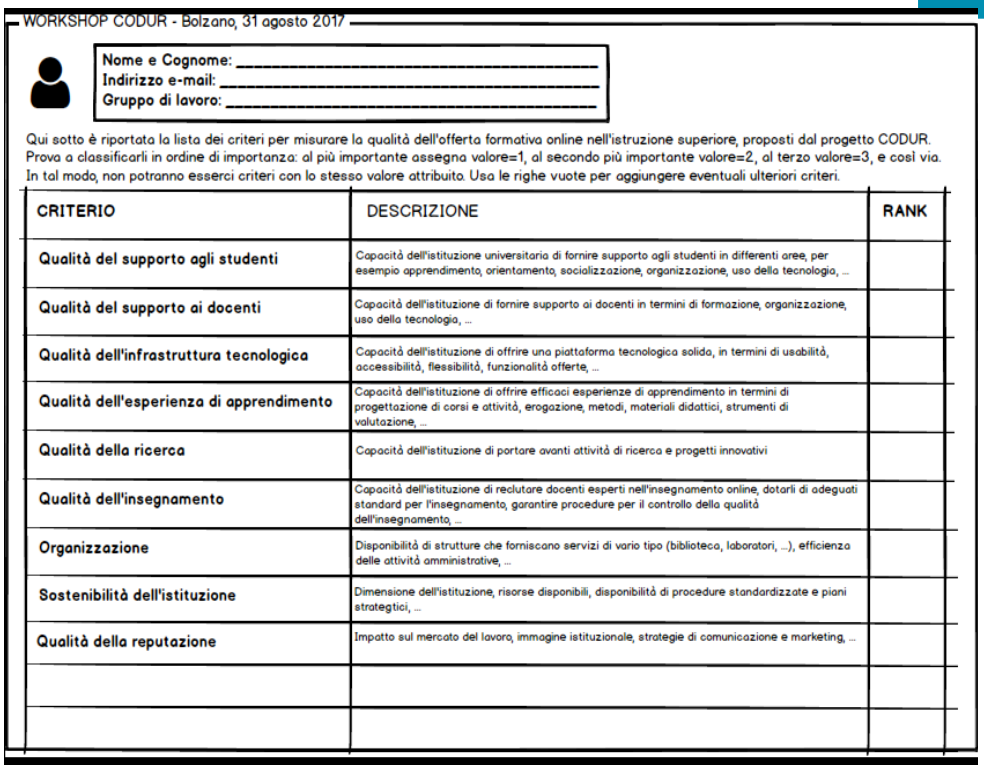
- A number of individual forms (one for each group member) to provide an individual ranking of the proposed criteria (see Figure 9).
- One collective form for the group to provide a collective ranking of the proposed criteria (see Figure 10).

Each group was asked to proceed as it follows:

1) Each person was asked to individually rank the proposed criteria (from 1 to 9, it was required to assign values reflecting the perceived importance of each criterion).
2) The group was asked to discuss the individual rankings and achieve an agreement on a common ranking, to be reported in the collective form.

It should be noted that a couple of critical aspects emerged during the activity that needed to be fixed:

- Both groups felt the need to clarify whether the required ranking should address online Universities or traditional University offering also online courses / programmes. For the CODUR project sake, we asked the groups to focus on online Universities.
- It was also necessary to stress the fact that the criteria and indicators proposed by CODUR are intended to address an online Higher Education *institution*, rather than courses or programmes.



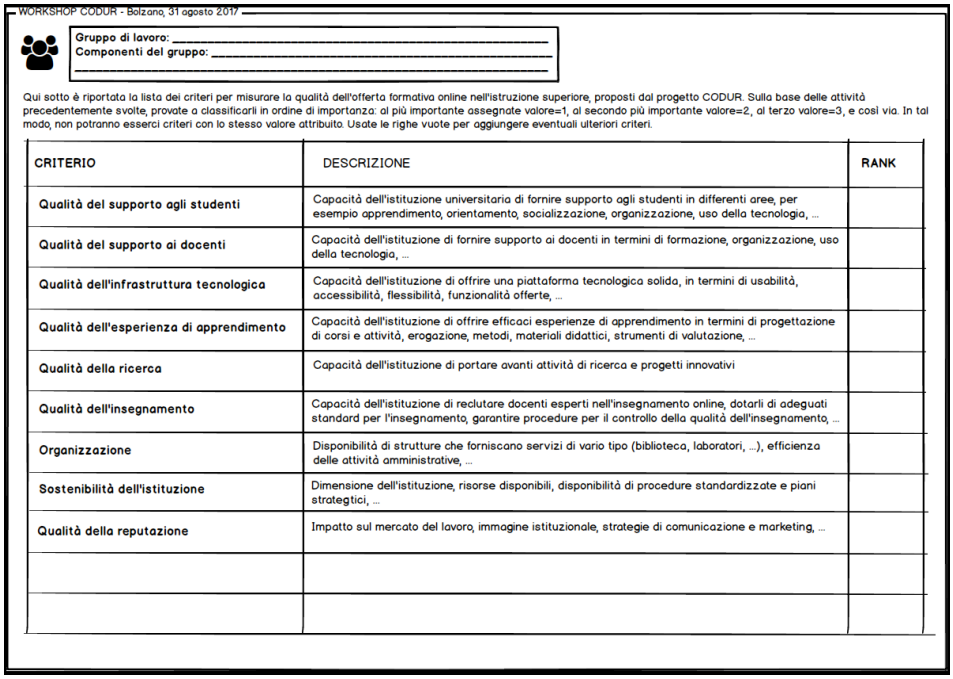
WORKSHOP CODUR - Bolzano, 31 agosto 2017

Nome e Cognome: _____
Indirizzo e-mail: _____
Gruppo di lavoro: _____

Qui sotto è riportata la lista dei criteri per misurare la qualità dell'offerta formativa online nell'istruzione superiore, proposti dal progetto CODUR. Prova a classificarli in ordine di importanza: al più importante assegna valore=1, al secondo più importante valore=2, al terzo valore=3, e così via. In tal modo, non potranno esserci criteri con lo stesso valore attribuito. Usa le righe vuote per aggiungere eventuali ulteriori criteri.

CRITERIO	DESCRIZIONE	RANK
Qualità del supporto agli studenti	Capacità dell'istituzione universitaria di fornire supporto agli studenti in differenti aree, per esempio apprendimento, orientamento, socializzazione, organizzazione, uso della tecnologia, ...	
Qualità del supporto ai docenti	Capacità dell'istituzione di fornire supporto ai docenti in termini di formazione, organizzazione, uso della tecnologia, ...	
Qualità dell'infrastruttura tecnologica	Capacità dell'istituzione di offrire una piattaforma tecnologica solida, in termini di usabilità, accessibilità, flessibilità, funzionalità offerte, ...	
Qualità dell'esperienza di apprendimento	Capacità dell'istituzione di offrire efficaci esperienze di apprendimento in termini di progettazione di corsi e attività, erogazione, metodi, materiali didattici, strumenti di valutazione, ...	
Qualità della ricerca	Capacità dell'istituzione di portare avanti attività di ricerca e progetti innovativi	
Qualità dell'insegnamento	Capacità dell'istituzione di reclutare docenti esperti nell'insegnamento online, dotarli di adeguati standard per l'insegnamento, garantire procedure per il controllo della qualità dell'insegnamento, ...	
Organizzazione	Disponibilità di strutture che forniscano servizi di vario tipo (biblioteca, laboratori, ...), efficienza delle attività amministrative, ...	
Sostenibilità dell'istituzione	Dimensione dell'istituzione, risorse disponibili, disponibilità di procedure standardizzate e piani strategtici, ...	
Qualità della reputazione	Impatto sul mercato del lavoro, immagine istituzionale, strategie di comunicazione e marketing, ...	

Fig. 9 – Form for individual ranking



WORKSHOP CODUR - Bolzano, 31 agosto 2017

Gruppo di lavoro: _____
Componenti del gruppo: _____
_____

Qui sotto è riportata la lista dei criteri per misurare la qualità dell'offerta formativa online nell'istruzione superiore, proposti dal progetto CODUR. Sulla base delle attività precedentemente svolte, provate a classificarli in ordine di importanza: al più importante assegnate valore=1, al secondo più importante valore=2, al terzo valore=3, e così via. In tal modo, non potranno esserci criteri con lo stesso valore attribuito. Usate le righe vuote per aggiungere eventuali ulteriori criteri.

CRITERIO	DESCRIZIONE	RANK
Qualità del supporto agli studenti	Capacità dell'istituzione universitaria di fornire supporto agli studenti in differenti aree, per esempio apprendimento, orientamento, socializzazione, organizzazione, uso della tecnologia, ...	
Qualità del supporto ai docenti	Capacità dell'istituzione di fornire supporto ai docenti in termini di formazione, organizzazione, uso della tecnologia, ...	
Qualità dell'infrastruttura tecnologica	Capacità dell'istituzione di offrire una piattaforma tecnologica solida, in termini di usabilità, accessibilità, flessibilità, funzionalità offerte, ...	
Qualità dell'esperienza di apprendimento	Capacità dell'istituzione di offrire efficaci esperienze di apprendimento in termini di progettazione di corsi e attività, erogazione, metodi, materiali didattici, strumenti di valutazione, ...	
Qualità della ricerca	Capacità dell'istituzione di portare avanti attività di ricerca e progetti innovativi	
Qualità dell'insegnamento	Capacità dell'istituzione di reclutare docenti esperti nell'insegnamento online, dotarli di adeguati standard per l'insegnamento, garantire procedure per il controllo della qualità dell'insegnamento, ...	
Organizzazione	Disponibilità di strutture che forniscano servizi di vario tipo (biblioteca, laboratori, ...), efficienza delle attività amministrative, ...	
Sostenibilità dell'istituzione	Dimensione dell'istituzione, risorse disponibili, disponibilità di procedure standardizzate e piani strategtici, ...	
Qualità della reputazione	Impatto sul mercato del lavoro, immagine istituzionale, strategie di comunicazione e marketing, ...	

Fig. 10 – Form for collective ranking

While the individual ranking phase was carried out by all the participants quite easily, the collaborative phase was more challenging.

In particular, one group had great difficulties to find an agreement about the ranking and in the end, even if in the final plenary they reported a number of very interesting reflections and contributions, they were not able to produce a unique ranking (see Section 4.2.3 for more details about this).

The other group, instead, in order to achieve a collective ranking, applied an algorithm in such a way to 'force' the consensus (see Section 4.2.3 for more details about this).

In both cases, the provided contribution is quite interesting and all the opinions and ideas gathered can nurture the work done so far by the project, as it is further illustrated in Section 4.2.3.

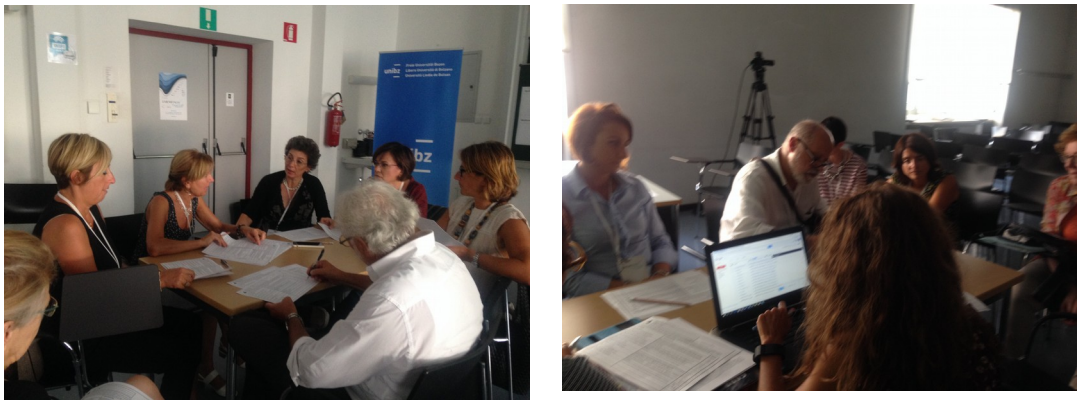


Fig. 11 – The working groups

4.2.3 Output

This section reports considerations emerging from the EMEM workshop. The first part regards the outputs from the Round Table (morning session), while the second one the outputs from the collaborative activity (afternoon session).

In order to explore the data collected during the workshop, a process of analysis was carried out. The data from the Round Table were analyzed following a thematic analysis approach (Braun & Clarke, 2006), while the data from the collaborative activity were tackled by means of a quali-quantitative approach.

Round Table (morning)

Through the discussion among the six experts participating in the Round Table, some interesting themes emerged with regard to the proposed topic ("Towards the recognition of the e-learning dimension in the University ranking systems"). The key questions asked to the experts as an input to their contributions included:

- What is your position with regard to university rankings?
- Do you think that e-learning should be considered in university ranking systems?
- What are your reflections/considerations regarding CODUR's suggested criteria and indicators?

The Round Table discussion was recorded and is currently available at the following url: .

The recorded session of the Round Table was repeatedly watched and analyzed to identify pre-determined and emerging themes and patterns deductively. Pre-determined themes were chosen based on the other main tasks of the CODUR project (i.e. IO1-A1 and IO1-A2), while other themes emerged during the analysis of the video.
Overall, the focus was on the following set of themes:

- Epistemological aspects (definitions, peculiarities, field of applications)
    - Quality assurance
    - Accreditation

    - Ranking
- "*One size fits all*" approach in rankings design
- Levels to be considered (macro, such as institutional level; micro, such as course level)
- Data source providers in rankings
- Indicators and parameters for rankings (methodology)
    - Subjectivity of indicators
    - Transparency of indicators
    - Robustness of statistics.

Following is a summary of the interventions of the six experts.

EXPERT	SUMMARY
FALCINELLI	In her speech Prof. Falcinelli expresses a critical position against (ranking) systems, as – in her view - they have overcome the initial and prior goal that was to offer a quick consultation tool to guide students and families to choose the university. Today, she says, ranking systems tend to overlap evaluation systems, but the logic of the two is actually different. Evaluation systems have a formative value, offering a number of elements of information to be used to improve universities. This cannot be asked to ranking systems that are often vague as far as the underpinning indicators and the methodological notes that are often not made explicit. Rankings shows a result, but it is often unclear how such result has been achieved. Prof. Falcinelli expresses concerns about how rankings are built, how data are collected in order to attribute the ratings that allows to build the rankings. Furthermore, government-institutional sources are not always able to capture the complexity of the university system, but only considers what is registered at institutional level. In Prof. Falcinelli's view, the idea that rankings are not absolute, but relative, should be reaffirmed. It is necessary to understand the architecture that generates a ranking system with respect to another, the way the ranking is built, but also who is the promoter (i.e. CENSIS for Italy). Those in charge of evaluation and analysis systems know that a lot of attention has to be paid to indicators and the way they are chosen - so it very good that the CODUR workshop offers the opportunity of a participatory reflection on this issue. It is also important to take into account the validity and reliability of an indicator in relation to the criterion to be measured, to have the numeric calculation that allows to do the classification. Another critical aspect is that rankings tend to simplify a complex system (university). Overall, Prof. Falcinelli agrees that ranking systems should address the e-learning component. Since ranking systems are widespread, it is important to reflect on how the online courses provision should be considered in a way that it is not detached from the University system

	as a whole. In her view, first we should clarify the kind of online courses provision we are talking about, whether they are totally online or blended, etc. In any case, talking about online courses provision would require a wider reflection on university teaching as a whole, on its overall level of innovation, in terms of models and teaching strategies that put the student at the center of the learning process. As far as the criteria and indicators proposed by the CODUR project, Prof. Falcinelli says there is a need to integrate them. For instance, "Quality of teaching", "Quality of the learning experience" and "Quality of the technological infrastructure" should be considered in an integrated logic. There should be specific space for courses design and organization, the possibility of a broad and sustainable use of functional resources for design; the design/organization of a course requires a devoted team of specialists. Moreover, the indicators of the "Quality of learning experience" could be enriched with the following: ▪ presence of specific tools for the analysis of students' learning needs (for instance, in terms of diversity, inclusion); ▪ clear definition of the objectives expressed in terms of competences; ▪ organization of courses with respect to contents; ▪ quality of learning materials; ▪ integration with online resources (MOOCs, OERs); ▪ presence of tools for students to interact with materials, to build new products, to interact and communicate within social communities; ▪ tools for students' self-monitoring and self-evaluation (e.g. e-portfolio); ▪ tools for course personalization; ▪ in case of blended learning, integration with presence should be also considered.
GHISLANDI	Prof. Ghislandi starts her speech by acknowledging that in most of the famous ranking systems we find the same small set of universities in the first places. She then cites a couple of important studies that are critical against ranking systems: *"It is now proved that international rankings are a marketing tool, absolutely non-scientific rankings, whose purpose is to influence the opinion of students and families on which the best institutions in the world are. According to the criteria set by the publisher of the ranking. "*(De Nicolao, 2014) *"..... our view is that the Shanghai ranking, in spite of the media coverage it receives, does not qualify as a useful and pertinent tool to discuss the "quality" of academic institutions, let alone to guide the choice of students and family or to promote reforms of higher education systems ... "*(Billaut, Bouyssou and Vinck, 2009) Prof. Ghislandi focuses the attention on the reasons behind these criticisms. In particular, she mentions a number of criticalities:

	▪ Rankings tend to ignore important aspects, such as the amount of money invested by universities, the ratio teachers-students, etc. (*value for money*) ▪ There are issues related to the correctness of methodology and indicators, as one should pay attention to: o Nature and adequacy of indicators o Adequacy of the database used to evaluate scientific research o Correctness and robustness of the bibliometric indicators, etc. ▪ Lack of transparency - universities are often advised by the same bodies that then draw up the charts ▪ Unbalanced indicators related to research ▪ Over-valuation of natural and biomedical sciences with respect to humanistic disciplines ▪ Evaluation of universities rather than departments or institutes ▪ Inaccuracy. Moreover, rankings do not take into account how much students learn (teaching & learning) and aspects such as: ▪ Students graduated on time (bachelors & masters) ▪ Academic staff with doctorates ▪ Contacts with work environments (bachelors & master), etc. Given that measuring how much students learn is difficult, rankings tend to consider parameters that are more easily capturable. In any case, Prof. Ghislandi agrees we should consider e-learning in ranking systems, as they have got peculiarities that need to be addressed (blended or e-learning). As far as the criteria and indicators proposed by the CODUR project, Prof. Ghislandi suggests to: ▪ Clarify the level we are addressing: Institution, department or single course? ▪ Focus on: Quality of the technology infrastructure - Quality of the learning experience - Sustainability of the institution. ▪ As far as the Quality of the learning experience – it should be meant more in the sense of “Students Voice”, rather than from the teacher’s perspective (instructional design).
LIMONE	At the beginning of his speech, Prof. Limone provides an overview on the national context (Italy), where rankings are coming from the Italian agency (ANVUR). In particular, in Italy at the moment we have : ▪ The VQR system – it evaluates the scientific performance of universities and the quality of research in university departments. It is a rigid, stable and consolidated evaluation system. ▪ The AVA process (self-assessment, assessment, and accreditation) made by ANVUR; it is aimed to monitor the quality of higher education. This has not yet produced a ranking, but it will produce it soon. It encompasses quantitative indicators on each individual course of study, which enable comparison (per geographic area, scientific-disciplinary area, etc.). ▪ The TECO process (Evaluation of University Student Learning Outcomes); this has been experimented by ANVUR in two rounds.

	It is a standardized test on the general skills of students at the end of the course across the country. In this model, the product (outcome) is evaluated rather than the process. These processes are presented by Prof. Limone in a dynamic context where autonomy of Universities and control put in place by the central government are the two faces of the same coin in higher education quality regulation. This issue should be considered not only at a national level, but also at a wider one. Prof. Limone cites the European Association for Quality Assurance in Higher Education (ENQA) as the organization representing quality assurance and accreditation organisations from the European Higher Education Area and internationally. As far as e-learning is concerned, Prof. Limone points out at the moment there are several stakeholders working on the definition of possible indicators; among the others: ▪ The ENQA Working Group ▪ Several international associations (including those managing MOOCs) ▪ CRUI (Conferenza dei Rettori Italiana), the Italian Committee of University Rectors, which is also managing an observatory on MOOCs. Thanks to these and other initiatives, policies have been developed - or are being developed – that, if implemented, would produce a number of rankings to compare and evaluate the online dimension. Among the others, Prof. Limone reminds that also SIREM (Società Italiana di Ricerca sull'Educazione Mediale) proposes a list of indicators and criteria to consider the quality of online education/e-learning. Starting from the Italian legal framework, encompassing: ▪ DM 635 [General Guidelines for 2016-2018 University Programming and Indicators for Periodic Results Evaluation] ▪ DM 987 [Self Evaluation, Evaluation, Initial and Periodic Accreditation of Universities and Universities]. SIREM proposes a list of indicators as response to the current Italian legal framework (which is still considered unsatisfactory). The goal is not to develop a new ranking, but rather to identify quality assurance criteria that can trigger the continuous improvement of e-learning. Here follows the list of the SIREM criteria presented by Prof. Limone as a conclusion to his contribution and an input to the CODUR proposal: ▪ Teaching-learning and assessment model ▪ Curriculum design, course design and course delivery ▪ Materials and contents ▪ Digital learning environment ▪ Cooperation and interactivity ▪ Teachers (number, qualification, experience) ▪ Tutor and other support staff (number, qualification,

	experience) ▪ Learning analytics ▪ New compulsory professional figures ▪ New tools and technological services (video-lesson tools, video-conferencing tools, social tools) ▪ Continuous research on e-learning.
MINERVA	Prof. Minerva points out "accreditation" and "ranking" are two different processes. The accreditation system should lead to a binary solution (accreditation vs. non-accreditation). Prof. Minerva strongly believes in ranking systems, as he thinks it is right for a service user to have metrics helping him/her to evaluate the validity of the institution running the service. The problem is what criteria to adopt to this aim. Looking at the criteria proposed by both CODUR and SIREM (see synthesis of Prof. Limone's speech), there is a criticism in the proposed criteria, because they are all related to procedural aspects, rather than to the outcome. According to Prof.'s opinion, procedural aspects (i.e. how the institution works) has to do with accreditation, while ranking should rather look at the outcome of the process. If so, criteria and indicators for the online dimension should not be seen as something different from those measuring the institution as a whole. One should look in any case at the outcome that is the "graduated student", independently on whether s/he attended an online or a face-to-face course. Furthermore, according to Prof. Minerva, the present proposals of criteria by CODUR and SIREM are not convincing, as they both rely on assumptions on the best teaching models. But the problem is: who decides that a model is better than others? This could be done only by evaluating the achieved results. Overall, Prof. Minerva stresses the need to make an effort to keep the process of accreditation (which evaluates the process) separated from ranking, where a synthetic score should be given based on most objective data related to the outcome. Lastly, Prof. Minerva observes most of the CODUR criteria seem to be strongly correlated from a statistical viewpoint, while criteria should be independent on each other.
ROSSI	Prof. Rossi started his speech by pointing out rankings have got limitations, but are useful. Coming back to the evaluation processes run by ANVUR in Italy in the last years (see Limone's speech), Prof. Rossi says – even if there have been weaknesses – still the level of attention towards university teaching has changed as a result of these actions. Prof. Rossi agrees accreditation and ranking have two different roles. As far as rankings are concerned - he says - the main problem comes from the fact that in these systems qualitative measures are

	transformed into quantitative ones. As far as the opportunity to consider e-learning in ranking systems, there are considerations to be done. According to Prof. Rossi's view, the indicators proposed by CODUR are good for certain purposes, but not for others. For instance, such ranking could be useful to help a student interested in taking an online learning degree in choosing among different online universities; in this case, a ranking with specific indicators would help in comparing performances of different online universities. If, on the contrary, one student wants to compare online universities with face-to-face ones, this should *not* be done with the same set of indicators, one should use different sets of indicators depending on the purpose. We should take into account that presently in Italy we have got different scenarios, including 100% online universities, mixed online-face-to-face universities, traditional universities offering e-learning courses, etc. In any case, accreditation and ranking should be two separated processes, but it should be noted that some things are interwoven. For instance, with reference to traditional universities, the distinction proposed by ANVUR (which classifies universities on the basis of the amount of online activities being supplied) directly affects the course and the accreditation. It also affects the ability for a traditional University to offer elements of the online and then it can be very restrictive both with reference to the internal bureaucracies of the university and the characterization of the course provision. In the model of University which is emerging, the digital component is ever more important and the boundaries between traditional and online course provision are becoming more and more nuanced. All the elements are important for quality assessment (accreditation) and they should be used to generate rankings. However, the same elements may create problems with regard to accreditation processes. The two processes (accreditation and ranking) might be inconsistent, so – Prof. Rossi says - it is right to keep them separate but also to understand how they can affect each other. In this model of University, the elements provided online (to students) are increasingly important. It is important to include those elements in rankings in order to support the student in orientation. Additional indicators, beyond the teaching model, could include the presence of tutors, the possibility of online interactions, etc. Prof. Rossi concludes by pointing out relying on 'objective data' related exclusively to the outcome (see Minerva's speech) can be risky: for example, if we look at the percentages of employability of students when they have concluded their study at the university, this depends not only on the quality of the university itself, but also on the territory. This is one case of an apparently objective indicator, having aspects of subjectivity.
RUI	Prof. Rui agrees that ranking tells us something about the product, provided that criteria and indicators are chosen honestly and with competence. The indication is for the user who realizes that s/he can

	get a certain result. Another need is to evaluate the process, to find possible gaps or errors, with the aim to improve it. In this case, the focus is not on the outcome, but on the process, i.e. on internal procedures. Rankings give directions, and – according to Prof. Rui's point of view - it's fair that the online dimension is also included. Of course, online has got its own specificities, but it is not something separated from the rest. Then the approach should be to evaluate teaching and learning as a whole, and add elements (criteria) that are significant for online. Prof. Rui suggests to focus especially on the following fundamental criteria: ▪ Quality of student support – this needs to be done both for online and not online - mentoring and tutoring - interaction and communication with the student ▪ Quality of teacher support - teachers do not receive appropriate training for online teaching ▪ Quality of technology infrastructure - tools ▪ Quality of learning experience – it is very much linked with the Quality of student and teacher support and also with the Quality of teaching ▪ Quality of research – presence of e-learning centers devoted to research – there is a need to be extremely transversal. In order to involve teachers, an e-learning center that is mixed between a teaching center and a research center could help.

Collaborative activity (afternoon)

Participants to the workshop were asked to rank the nine criteria from most important (1) to least important (9). This task was analogous to the one performed by participants to the Delphi Study. However, since participants to the workshop worked with pen and paper, choosing the place in ranking for each criterion, ties were theoretically possible (whereas the Limesurvey procedure made ties impossible by design). In practice, only a single participant tied two criteria, ranking both reputation and sustainability as ninth by order of importance, so that no criterion occupied the eighth place. This tie was broken randomly (sustainability took the eighth spot).

At the end of the collaborative activity, some interesting feedback emerged from the groups. It should be noted that, of the 6 experts from the morning Round Table, 4 have taken full part in the afternoon activities, while 2 have participated in part.

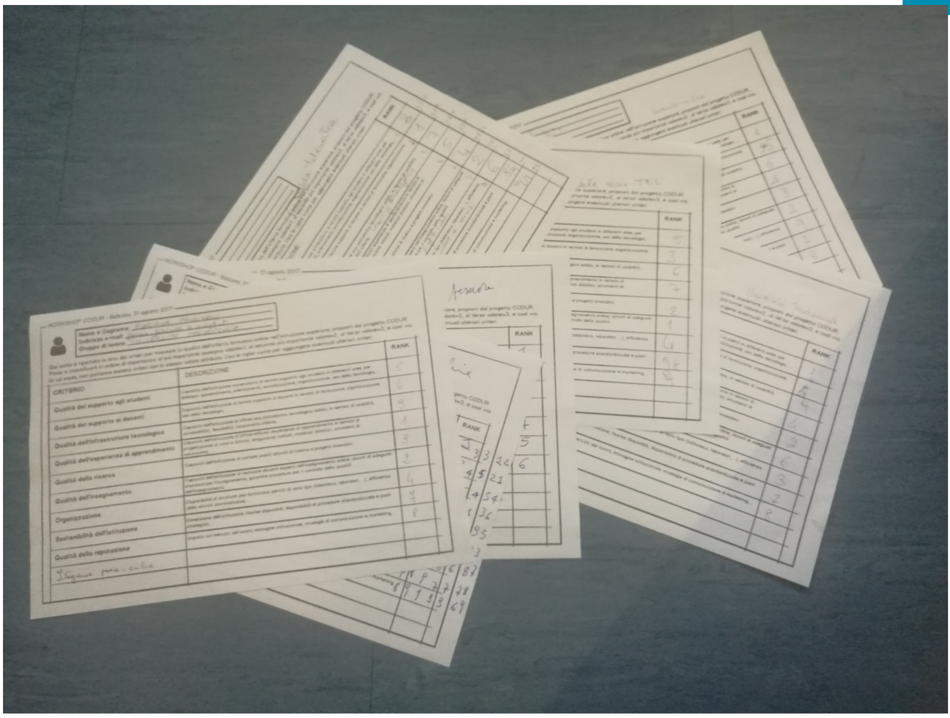
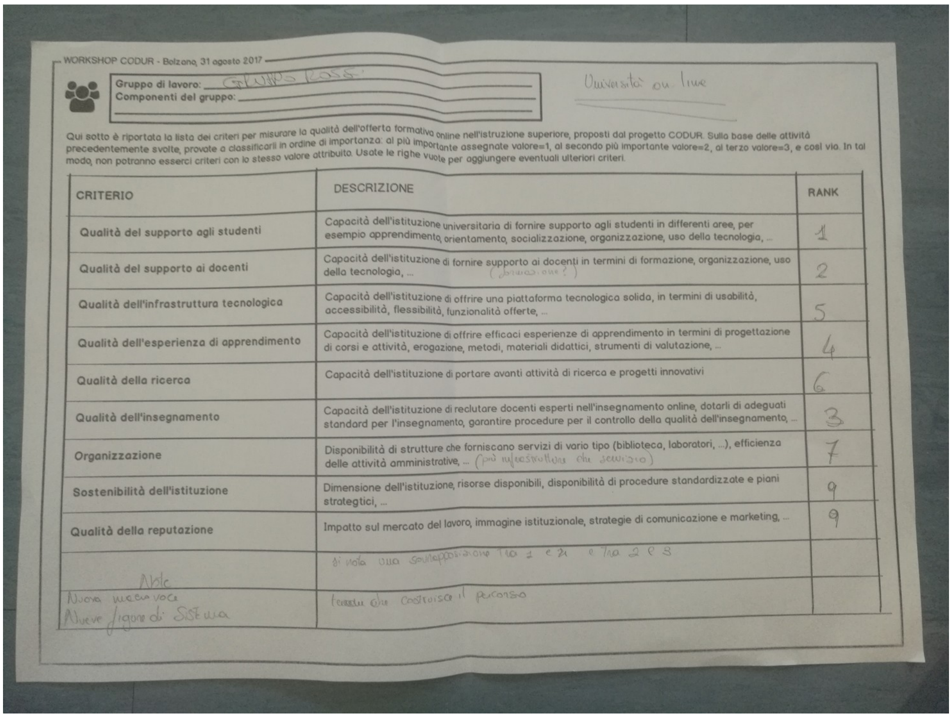


Fig. 12 – Individual ranking forms filled in



WORKSHOP CODUR - Bolzano, 31 agosto 2017

Gruppo di lavoro: Gruppo Rossi
Componenti del gruppo:

Università on line

Qui sotto è riportata la lista dei criteri per misurare la qualità dell'offerta formativa online nell'istruzione superiore, proposti dal progetto CODUR. Sulla base delle attività precedentemente svolte, provate a classificarli in ordine di importanza: al più importante assegnate valore=1, al secondo più importante valore=2, al terzo valore=3, e così via. In tal modo, non potranno esserci criteri con lo stesso valore attribuito. Usate le righe vuote per aggiungere eventuali ulteriori criteri.

CRITERIO	DESCRIZIONE	RANK
Qualità del supporto agli studenti	Capacità dell'istituzione universitaria di fornire supporto agli studenti in differenti aree, per esempio apprendimento, orientamento, socializzazione, organizzazione, uso della tecnologia, ...	1
Qualità del supporto ai docenti	Capacità dell'istituzione di fornire supporto ai docenti in termini di formazione, organizzazione, uso della tecnologia, ... (formazione?)	2
Qualità dell'infrastruttura tecnologica	Capacità dell'istituzione di offrire una piattaforma tecnologica solida, in termini di usabilità, accessibilità, flessibilità, funzionalità offerte, ...	5
Qualità dell'esperienza di apprendimento	Capacità dell'istituzione di offrire efficaci esperienze di apprendimento in termini di progettazione di corsi e attività, erogazione, metodi, materiali didattici, strumenti di valutazione, ...	4
Qualità della ricerca	Capacità dell'istituzione di portare avanti attività di ricerca e progetti innovativi	6
Qualità dell'insegnamento	Capacità dell'istituzione di reclutare docenti esperti nell'insegnamento online, dotarli di adeguati standard per l'insegnamento, garantire procedure per il controllo della qualità dell'insegnamento, ...	3
Organizzazione	Disponibilità di strutture che forniscano servizi di vario tipo (biblioteca, laboratori, ...), efficienza delle attività amministrative, ... (più infrastrutture che servizio)	7
Sostenibilità dell'istituzione	Dimensione dell'istituzione, risorse disponibili, disponibilità di procedure standardizzate e piani strategtici, ...	9
Qualità della reputazione	Impatto sul mercato del lavoro, immagine istituzionale, strategie di comunicazione e marketing, ...	9
	si nota una sovrapposizione tra 1 e 2 e tra 2 e 3	
Nuova voce Nuove figure di Sistema		

Fig. 13 – Collective form filled in by one of the group as a result of the group activity

Relative importance of criteria

The criteria rankings thus obtained were analyzed using the same procedure used for the Delphi Study. Figure 14 reports the importance estimates for this sample.

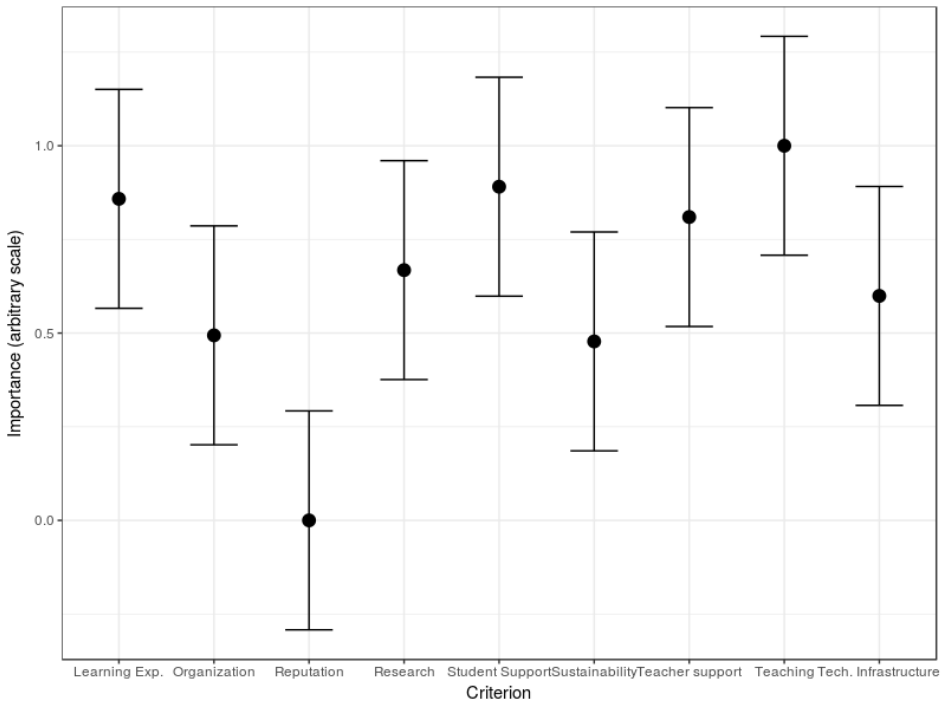


Fig. 14 - Relative importance of the evaluation criteria as estimated by workshop participants

Figure 15 compares the workshop estimates to the estimates resulting from the Delphi Study. In this figure, both scales have been rescaled to range from 0 to 1, in order to ease estimates comparison. Error bars for the workshop are much larger, since the sample size is smaller (N=10) than the Delphi Study's (N=40).

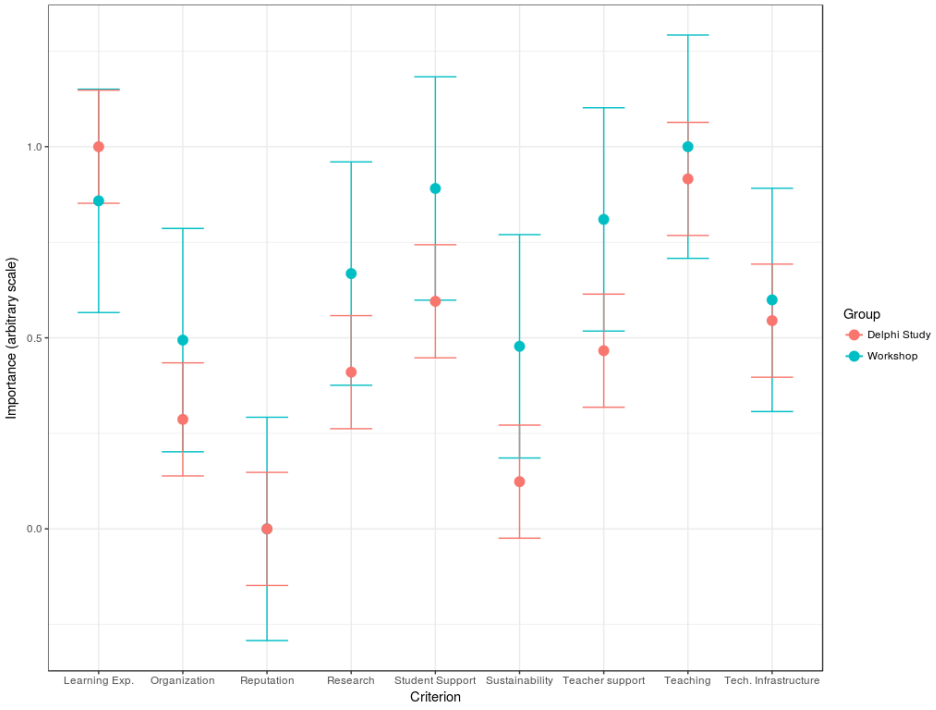


Fig. 15 - Relative importance of the evaluation criteria as estimated by participants to the Delphi Study (red) and to the workshop (blue).

By comparing the results of the two samples, we can see remarkable agreement, despite the different sample compositions and methodologies. Error bars for the two groups overlap for all criteria, and both groups agree that reputation is the least important of criteria. Largest - albeit non-significant - disagreements pertain to the perceived importance of sustainability of the institution and quality of student and teacher support: all three criteria were deemed more important by participants to the workshop.

By the analysis of the narratives emerged as qualitative feedback from the groups during the final plenary session, further reflections can be highlighted.

As general consideration, the groups agreed on the difficulty of keeping some of the proposed criteria separate and that the terms used to define the proposed criteria and parameters are in most of the cases subjectively interpretable, with a very wide range.

Following is a detailed summary of the different issues emerged.

- "Quality of student support" and "Quality of the learning experience" are considered overlapping but both very important.
- "Quality of teacher support" and "Quality of technology infrastructure" should consider aspects related to teachers' training as indicator(s).
- "Sustainability of the institution" and "Quality of reputation" are considered less important and it is not clear how to understand the first one ("Sustainability of the institution") rather than an appendix of the second "Quality of reputation".
- Some criteria should be added in order to consider specific system figures (design team, path design support, …).
- Parameters or indicators could be grouped into comprehensive categories (i.e. pedagogical, organizational, external relationships) would facilitate to put them in an order. In particular:
    - "Sustainability of the institution" and "Quality of reputation" are meaningful in terms of external image of the institution;
    - "Quality of student support", "Quality of teacher support", "Quality of the learning experience", and ""Quality of teaching" are all related to pedagogical aspects;
    - "Quality of technology infrastructure" and "Quality of organization" are the backbone of everything, without which any online institution could exist.
- "Quality of technology infrastructure" is a fundamental prerequisite, not only in terms of classical platforms, but also as *cloud* and *services*; it is anything that guarantees the workability of the online structure from the technological point of view (not just the platform).
- Some criteria are strictly related to the external image of the institution (see above). In particular, the criterion "Quality of reputation" was very important for some participants and not at all relevant for others. In the first case, it was meant as the impact on the labor market (result, product, in terms of people who are outsourced to work or are able to function in the work/occupation).

- "Quality of technology infrastructure" and "Quality of organization" are both considered important.
- "Quality of research" is considered not much relevant if research is meant as academic research (the one that aims to publish in international journals), while relevant if meant as something that has to do with reflection, monitoring of results, attention to innovation, ability to promote innovative projects, …
- In reference to pedagogical aspects, "Quality of teaching" and "Quality of teacher support" are considered related: if quality of teaching is guaranteed, then it means that behind there is a level of support for them (also in terms of training).

A more comprehensive comparison sheet about the CODUR criteria and indicators (with ranking and relative importance emerged both from the Delphi Study and the workshop) is presented in Appendix 3.

5 Discussion and main lessons learnt

In this section, we synthesize the main lessons learnt so far as a result of the actions carried out within IO1-A3.

There are general considerations and recommendations, as well as punctual feedback on the list of criteria proposed by CODUR.

In particular:

General lessons learnt

- Evaluation, accreditation and ranking are very critical aspects within the Higher Education community and – even if much has been said so far about these topics – we are far from being able to say a final word on them. In the ever changing and competitive world we live, being able to define indicators able to adequately measure the quality of Universities is still perceived as fundamental. However, even if evaluation, accreditation and ranking are somehow all sides of the same coin, it is important not to mix these terms up, as they point to different actions, each one with different aims. In CODUR we have decided to focus on the ranking area.
- The lack of specific indicators for measuring the quality of online learning, is definitely felt as an urgent gap to be filled in by the HE community. In this sense, CODUR has identified a real need, i.e. to define specific criteria and indicators for the online dimension.
- When we use the term "online dimension" within the HE context, we can point to many different situations, ranging from 'completely online institutions' (as the Open Universities), to traditional Universities running only a few courses or entire programmes through the Internet. Defining criteria and indicators for these situations (or any possible variant) is something extremely delicate and one should choose the exact focus of the work. In CODUR, we have chosen to focus on the evaluation of *online Higher Education institutions*, rather than on courses or programmes. On the other hand, it has been very useful for the consortium to gather opinions from experts with a solid background in traditional Universities, but at the same time highly competent on the topics of online education. In fact, those experts are able to inform the project with their viewpoints on a very complex and dynamic reality and to contribute to the ongoing discussion on both aspects.
- Another aspect that has clearly emerged from the work done so far is that existing ranking systems are controversial (Amsler & Bolsmann, 2012; Barron, 2017; Çakır et al., 2015): even if some recognize the need to compare and classify Universities, others claim ranking systems are not academic tools, but rather marketing ones. Among the main weaknesses often mentioned for the existing ranking systems, many say they are not solid enough, especially as far as validity of indicators, methodological correctness, transparency of sources of information and algorithms, etc. (Billaut et al., 2009).

Feedback on the CODUR proposed criteria and indicators

Overall, taking on board the feedback collected so far thanks to the Delphi Study and the EMEM workshop, we can conclude the present list of criteria and indicators is a valuable piece of work and a good starting point to determine a set of criteria and indicators for online HE institutions. In the following, we synthesise the main suggestions collected so far

that should be used in the future by the project to further improve the existing set of criteria:

- (Some of) the present indicators seem to refer to the institution as a whole (independently on the online dimension), while others are specific for the online dimension. This might derive from the fact that within the project it is not yet clear enough whether the final output will be integrated within one existing ranking system (such as for example U-Multirank), or whether it will stay as a stand-alone set of indicators. Consequently, sometimes indicators that in principle should already been present in any existing system, have been also reported in the CODUR set, but these should definitely be deleted and not replicated once integrated in an existing system.
- Present indicators are not 'actual indicators', as most of them are not straightforwardly operationalizable. Besides, some indicators are quite high, others are more fine-grained. In the future, the project should try to make them more homogenous, by choosing the level to focus on.
- Furthermore, it is not clear whether we are looking for a quantitative or qualitative indicators. Presently, indicators are mixed and even if a mixed approach is in principle not negative, this should be clearly chosen and stated.
- According to the feedback collected so far, "Quality of learning experience" and "Quality of teaching" (even if very much related) should be merged and considered the most important criteria. As far as their indicators are concerned, CODUR could consider adding indicators such as: availability of analysis for students' needs definition, availability of design teams (composed of people with various competences), availability of self-monitoring and evaluation means & portfolios, etc.
- In order to evaluate the "Quality of learning experience" and the "Quality of teaching", greater importance should be given to the *distance travelled*, rather than to the *destination*: how much do students learn? What competences have been gained by students as a result of the programme?
- Some of the present criteria are very much intertwined and the boundaries between them is often blurred. For example, according to our results so far, "Quality of student support" and "Quality of learning experience" are probably very much related; the same can be said for "Quality of teacher support" and "Quality of teaching". A statistical correlation would be interesting to investigate this issue. More in general, some of the interviewed experts raised the objection that the criteria should be "independent variables". Though, if we accept this objection, it may become very difficult to find two independent criteria. How could CODUR solve the dilemma?
- "Quality of reputation" and "Sustainability of the institution" have been positioned in the last places, but this might come from the fact that they are somehow a result of all the other criteria.
- The present indicators for "Quality of research" are very much oriented to publications, rather than reflection.

6 Conclusions

In this document, we have reported the work done in the CODUR project under Task IO1-A3. According to the project proposal, main aim of this task is to develop, test and refine representative performance online quality education indicators based on common criteria.

The document describes the overall methodology adopted in Task IO1-A3, the preliminary proposal of the CODUR criteria and indicators, along with the process that led the consortium to its preliminary definition. Then the document contains the description of the two actions used to collect external feedback and reactions to the CODUR criteria and indicators, i.e. the Delphi Study and the EMEM workshop. Both the actions have been described in this document and their respective outputs have been reported. On the top of this, we have discussed the main lessons learnt and have provided recommendations for future work.

Overall, the actions put in place so far by the consortium have turned out to be quite effective in terms of feedback collected.

According to the overall methodology adopted in this task (see Fig. 1) such feedback will be integrated into the original CODUR proposal and a revised and enhanced version of the list of criteria and indicators will be issued in the next months of the project. This will be then further refined and tuned thanks to Round 2 of the Delphi Study, which is expected to be carried out between Autumn 2017 and early Winter 2018.

The work will also continue with IO2, which envisages the definition of Guidelines for integrating online education quality assurance metrics in existing ranking systems, such as for example, the U-Multirank. This should prove the feasibility of enriching an existing ranking system with ad hoc criteria and indicators specifically addressing the online dimension of HE institutions.

7 References

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

Barron, G. R. S. (2017). The Berlin Principles on Ranking Higher Education Institutions: limitations, legitimacy, and value conflict. *Higher Education*, *73*(2), 317-333. doi: 10.1007/s10734-016-0022-z

Billaut, J. C., Bouyssou, D., & Vincke, P. (2009). Should you believe in the Shanghai ranking?. 43 pages. Retrieved from: https://hal.archives-ouvertes.fr/file/index/docid/388319/filename/Shanghai_JCB_DB_PV.pdf

Brasher, A., Holmes, W. & Whitelock, D. (2017). A means for systemic comparisons of current online education quality assurance tools and systems. CODUR Project Deliverable IO1.A2. Retrieved from: https://drive.google.com/drive/folders/0B433yeYVgtlaa1dlVUpfaEtoSHM

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative research in psychology*, *3*(2), 77-101.

Çakır, M. P., Acartürk, C., Alaşehir, O., & Çilingir, C. (2015). A comparative analysis of global and national university ranking systems. *Scientometrics*, *103*(3), 813-848.

Dalkey, N., & Helmer, O. (1963). An experimental application of the Delphi method to the use of experts. *Management science*, *9*(3), 458-467.

Linstone, H. A., & Turoff, M. (2002). The Delphi Method. *Techniques and applications*, *53*. Retrieved from:

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

Marasini, D. Quatto, P. Ripamonti, E. (2014). Assessing the inter-rater agreement for ordinal data through weighted indexes. Statistical methods in medical research.

McAleese, M., Bladh, A., Bode, C., Muehlfeit, J., Berger, V., & Petrin, T. (2014). Report to the European Commission on new modes of learning and teaching in Higher Education.

Ossiannilsson, E., Williams, K., Camilleri, A. F., & Brown, M. (2015). Quality Models in Online and Open Education around the Globe: State of the Art and Recommendations. Oslo: International Council for Open and Distance Education (ICDE).

Plesch C., Kaendler C., Deiglmayr A., Mullins D., Rummel N., Spada H. (2012). Lo studio Delphi di STELLAR sul Technology Enhanced Learning. *TD Tecnologie Didattiche*, 20 (3), pp. 144-154

Rowe G., Wright G. (1999). The Delphi technique as a forecasting tool: issues and analysis.

Thurstone, L. L. (1927) A law of comparative judgments. Psychological Review, 34, 273-286.

Van Vught, F., & Ziegele, F. (2011). Design and testing the feasibility of a multidimensional global university ranking. *Final Report. CHERPA-Network*.

Appendix 1. Questionnaire Round 1 of the Delphi study

CODUR - Towards the creation of an online dimension for University Rankings

(http://in3.uoc.edu/opencms_in3/opencms/webs/projectes/codur/en/index.html)

Dear participant,
Thank you for your agreeing to participate in the CODUR Delphi Study.

The aim of this study is to ask for views on ***indicators for evaluating the quality of online higher education institutions***. The study will compare views within a selected group of interested stakeholders (including students, teachers and researchers). For the purpose of this study, we are defining "online higher education institutions" as those that offer degrees.

In this survey you will be presented with a variety of ***criteria*** and ***indicators*** for university raking systems, clustered in nine themes. You will be asked to evaluate the importance of these criteria and indicators based on your professional experience.

The themes are:

- Learning experience,
- Teaching,
- Research,
- Organization of the institution,
- Sustainability of the institution,
- Reputation of the institution,
- Student learning and (social) 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 27 questions in this survey

Some questions about you

[]Gender

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

Female Female

Male 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 student

...a researcher ...a researcher

...an educator (higher education) ...an educator (higher education)

...an educator (school) ...an educator (school)

...an educator (workplace) ...an educator (workplace)

...a policy maker ...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 Not at all informed

Slightly informed Slightly informed

Well informed Well informed

Very well informed Very well informed

[]
If you are willing to be contacted about this study, please provide your e-mail address below. This is entirely voluntary/optional

Please write your answer here:

Quality of student support

Please rate the importance of the following aspects related to the **quality of student support.**

Criterion definition: *Quality of student support refers to the ability of the online higher education institution to provide support to learners in many different areas, for example learning, orientation, socializing with peers, organisational issues, use of technology, etc.*

Please indicate how important you consider each aspect to be, from 0 (not at all important) to 4 (extremely important). Please give a response for EVERY aspect and rate each aspect *independently* of the importance of the overall criterion.

[] *

Please choose the appropriate response for each item:

	0 (Not at all important)	1	2	3	4 (Extremely important)	Unsure / unclear
Orientation services to help learners taking decisions about their learning path						
Mentoring, tutoring activities to support student learning						
Support to student community building						
Support to alumni community building						
FAQ systems, Helpdesk, or sim. for learners						
Newsletters or sim. to allow communication from the institution to the learner						
Regularity of information update						
Technology support						

[]Are there any other aspects for evaluating the **quality of student support** that you think are important to consider?

Please write your answer here:

Quality of teacher support

Please rate the importance of the following aspects related to the **quality of teacher support.**

Criterion definition: *Quality of teacher support refers to the ability of the online higher education iinstitution to provide support to teachers and lectures in terms of training provision, organisational issues, use of technology, etc.*

Please indicate how important you consider each aspect to be, from 0 (not at all important) to 4 (extremely important). Please give a response for EVERY aspect and rate each aspect *independently* of the importance of the overall criterion.

[] *

Please choose the appropriate response for each item:

	0 (Not at all important)	1	2	3	4 (Extremely important)	Unsure / unclear
Opportunities for teaching staff to be trained in online education						
Support to teachers and lecturers (for example provided via a specific department devoted to this)						
Support to teacher community building						
FAQ systems, Helpdesk or sim. for teachers						

	0 (Not at all important)	1	2	3	4 (Extremely important)	Unsure / unclear
Newsletters or sim. to allow communication from the institution to the teaching staff						
Regularity of info update						
Technology support						

[]Are there any other aspects for evaluating the **quality of teacher support** that you think are important to consider?

Please write your answer here:

Quality of technology infrastructure

Please rate the importance of the following aspects related to the **quality of technological infrastructure.**

Criterion definition: *Quality of the technological infrastructure refers to the ability of the online higher education iinstitution to offer a sound technological platform, in terms of usability, accessibility, flexibility, types of features offered, etc.*

Please indicate how important you consider each aspect to be, from 0 (not at all important) to 4 (extremely important). Please give a response for EVERY aspect and rate each aspect *independently* of the importance of the overall criterion.

[] *

Please choose the appropriate response for each item:

	0 (Not at all important)	1	2	3	4 (Extremely important)	Unsure / unclear
Compliance of interfaces with usability and accessibility standards						

	0 (Not at all important)	1	2	3	4 (Extremely important)	Unsure / unclear
Consistency/ robustness of the platform						
Capacity and concurrence of the platform (i.e. number of people that the infrastructure can host, even at the same time)						
Flexibility and scalability of the platform in view of future changes /new emerging needs						
Capability of the platform to satisfy different user needs (personalization capabilities, etc.)						
Capability to support interaction and teamwork						
Capability to support evaluation						
Adequacy of the technical support (provided by department/ ad hoc team/or external company in charge of the technology)						
Adequacy of data security mechanisms						
Existence of a plan for system maintenance and contingency management						

[]Are there any other aspects for evaluating the **quality of technological infrastructure** that you think are important to consider?

Please write your answer here:

Quality of the learning experience

Please rate the importance of the following aspects related to the **quality of the learning experience.**

Criterion definition: *Quality of the learning experience refers to the ability of the online higher education iinstitution to offer effective learning experiences, in terms of sound design, delivery, adopted methods, learning materials, assessment means, etc.*

Please indicate how important you consider each aspect to be, from 0 (not at all important) to 4 (extremely important). Please give a response for EVERY aspect and rate each aspect *independently* of the importance of the overall criterion.

[] *

Please choose the appropriate response for each item:

	0 (Not at all important)	1	2	3	4 (Extremely important)	Unsure / unclear
Quality of course / learning design						
Quality of pedagogy /methodology						
Quality of assessment						
Quality of learning materials						

[]Are there any other aspects for evaluating the **quality of the learning experience** that you think are important to consider?

Please write your answer here:

Quality of research

Please rate the importance of the following aspects related to the **quality of research.**

Criterion definition: *Quality of research refers to the ability of the online higher education institution to carry out research initiatives and innovation projects*

Please indicate how important you consider each aspect to be, from 0 (not at all important) to 4 (extremely important). Please give a response for EVERY aspect and rate each aspect *independently* of the importance of the overall criterion.

[] *

Please choose the appropriate response for each item:

	0 (Not at all important)	1	2	3	4 (Extremely important)	Unsure / unclear
Research output (in terms of publications, visiting scholars, ...)						
Research in online teaching & learning (research groups, research projects, etc.)						
Teaching staff engaged in research in online education						
Internal centres devoted to research in online education						
Research in other disciplines						
Use of research for improvement and innovation (research based projects…)						

[]Are there any other aspects for evaluating the **quality of research** that you think are important to consider?

Please write your answer here:

Quality of teaching

Please rate the importance of following aspects related to the **quality of teaching.**

Criterion definition: *Quality of teaching refers to the ability of the online higher education iinstitution to recruit experienced teachers trained in delivering online teaching, provide them with standards for teaching, guarantee regular quality control procedures, etc.*

Please indicate how important you consider each aspect to be, from 0 (not at all important) to 4 (extremely important). Please give a response for EVERY aspect and rate each aspect *independently* of the importance of the overall criterion.

[] *

Please choose the appropriate response for each item:

	0 (Not at all important)	1	2	3	4 (Extremely important)	Unsure / unclear
Profile of the teaching staff						
Teacher assessment and quality control						
Involvement of academics in teaching						
Standards for regulating teacher-student interactions (teachers' response time, modalities, timeliness of feedback provision for tests, etc.)						

[]Are there any other aspects for evaluating the **quality of teaching** that you think are important to consider?

Please write your answer here:

Quality of organization

Please rate the importance of the following aspects related to the **quality of organization.**

Criterion definition: *Quality of organization includes aspects such as availability of structures providing services of various kind, lightness and efficiency of bureaucracy, etc.*

Please indicate how important you consider each aspect to be, from 0 (not at all important) to 4 (extremely important). Please give a response for EVERY aspect and rate each aspect *independently* of the importance of the overall criterion.

[] *

Please choose the appropriate response for each item:

	0 (Not at all important)	1	2	3	4 (Extremely important)	Unsure / unclear
Decentralized structures on the territory						
Structures such as libraries, labs,. etc.						
Non instructional support services (providing assistance for admission, financial issues, registration, enrollment,etc.)						
Examination policies able to cater with for the needs of e-learning courses						
Credit transfer system aligned with national (and /or European) systems and operates bidirectionally						
Existence of a complaints and appeals system for learners						

[]Are there any other aspects for evaluating the **quality of organization** that you think are important to consider?

Please write your answer here:

Sustainability of the institution

Please rate the importance of the following aspects related to the **sustainability of the institution.**

Criterion definition: *Sustainability includes aspects such as the size of the institution, resources, availability of standardised procedures and strategic plans, etc.*

Please indicate how important you consider each aspect to be, from 0 (not at all important) to 4 (extremely important). Please give a response for EVERY aspect and rate each aspect *independently* of the importance of the overall criterion.

[] *

Please choose the appropriate response for each item:

	0 (Not at all important)	1	2	3	4 (Extremely important)	Unsure / unclear
Institutional strategic plan for online education						
Size of the institution						
Overall coherence of program design and provision (interconnections among courses, flexibility of the design, clarity of program design, …)						

	0 (Not at all important)	1	2	3	4 (Extremely important)	Unsure / unclear
Existence of standardized workflows for program /course /material design and development						
Sustainability of the portfolio of programmes						
Resources (including financial ones) specifically devoted to the online program						
Clear policy regarding OERs and MOOCs						

[]Are there any other aspects for evaluating the **sustainability of the institution** that you think are important to consider?

Please write your answer here:

Quality of reputation

Please rate the importance of the following aspects related to the **"reputation" or impact of the institution.**

Criterion definition: *Reputation includes aspects such as impact on job market, institutional image, communication strategies, etc.*

Please indicate how important you consider each aspect to be, from 0 (not at all important) to 4 (extremely important).
Please give a response for EVERY aspect and rate each aspect *independently* of the importance of the overall criterion.

[] *

Please choose the appropriate response for each item:

	0 (Not at all important)	1	2	3	4 (Extremely important)	Unsure / unclear
Job opportunities for graduates						
Internship and mobility opportunities						
Social impact						
Institutional image						
Communication strategies						
Organization of events, conferences, etc.						

[]Are there any other aspects for evaluating the **"reputation" or impact of the institution** that you think are important to consider?

Please write your answer here:

Ranking criteria to evaluate online higher education institutions

Below you will find the complete list of the criteria proposed in this survey to evaluate online higher education institutions.
Please rank them by order of importance (with the most important aspect on top, and the least important aspect at the bottom)

[]
We define the criteria as it follows:

Quality of student support - it refers to the ability of the online higher education institution to provide support to learners in many different areas, for example learning, orientation, socializing with peers, organisational issues, use of technology, etc.

Quality of teacher support - it refers to the ability of the online higher education institution to provide support to teachers and lectures in terms of training provision, organisational issues, use of technology, etc.

Quality of technological infrastructure - it refers to the ability of the online higher education institution to offer a sound technological platform, in terms of usability, accessibility, flexibility, types of features offered, etc.

Quality of the learning experience - it refers to the ability of the online higher education institution to offer effective learning experiences, in terms of design, delivery, methods, learning materials, assessment means, etc.

Quality of research - it refers to the ability of the online higher education institution to carry out research initiatives and innovation projects

Quality of teaching - it refers to the ability of the online higher education institution to recruit experienced teachers trained in delivering online teaching, provide them with standards for teaching, guarantee regular quality control procedures, etc.

Organization - it includes aspects such as availability of structures providing services of various kind, lightness and efficiency of bureaucracy, etc.

Sustainability of the institution - it includes aspects such as the size of the institution, resources, availability of standardised procedures and strategic plans, etc.

"Reputation" or impact - it includes aspects such as impact on the job market, institutional image, communication strategies, etc.

*

All your answers must be different and you must rank in order.

Please number each box in order of preference from 1 to 9

Quality of student learning and (social) support	
Quality of teacher support	

Quality of technological infrastructure

Quality of the learning experience

Quality of research

Quality of teaching

Organization

Sustainability of the institution

"Reputation" / Impact

[]

Are there any other aspects for evaluating the quality of online higher education institutions that you believe should be considered (in addition to those listed above)?

Please write your answer here:

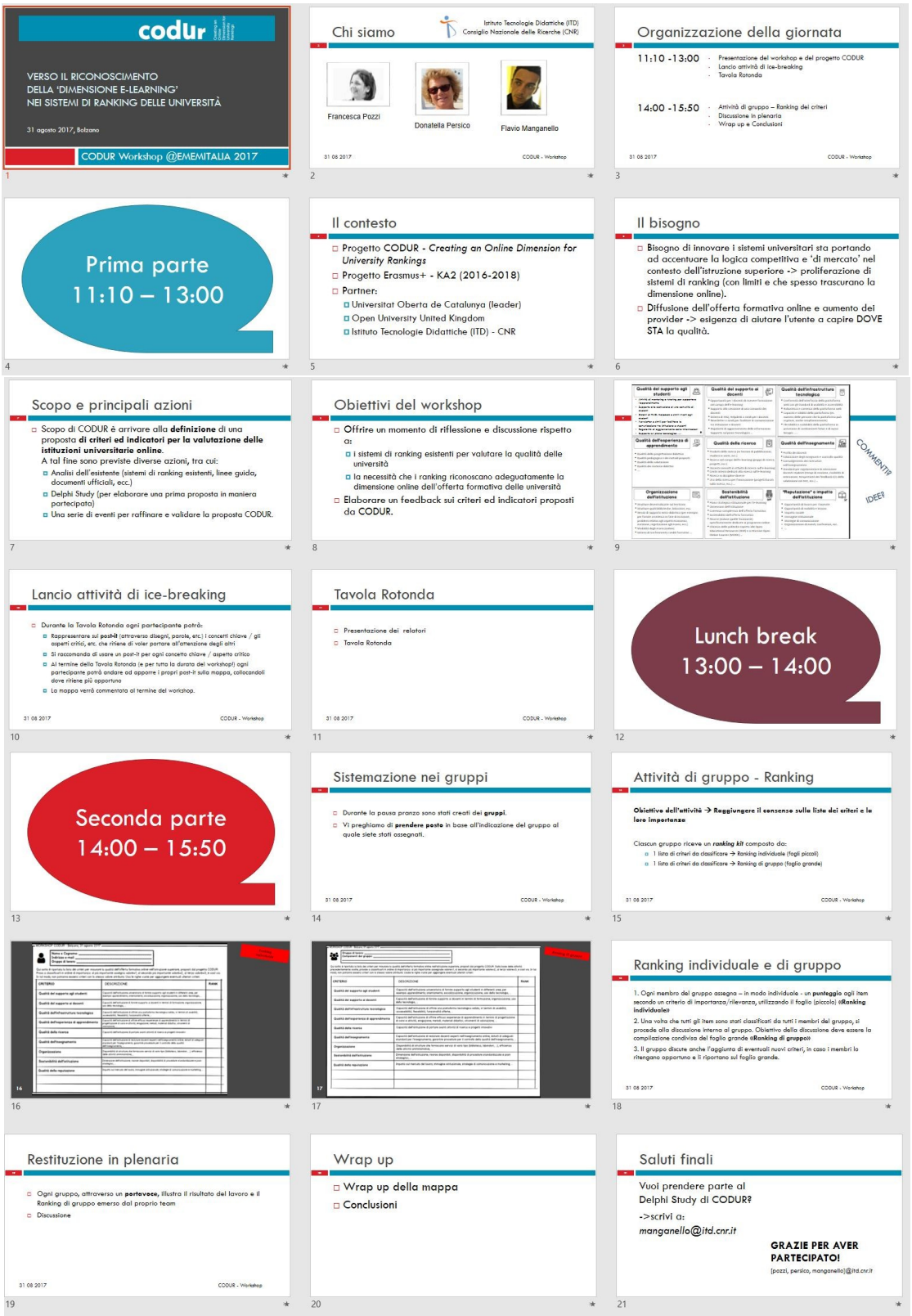
Any other comments

[]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.

Appendix 2. Presentation held at the EMEM workshop



Pozzi et al., 2017

Appendix 3. CODUR Criteria and Indicators – Rank and relative importance

RANK	CRITERIA (relative importance - DELPHI)	CRITERIA (relative importance - EMEM Workshop)		
1	Quality of learning experience	Quality of teaching		
2	Quality of teaching	Quality of student support		
3	Quality of student support	Quality of learning experience		Colors legend
4	Quality of technological infrastructure	Quality of teacher support		PRODUCTS
5	Quality of teacher support	Quality of research		MANAGEMENT
6	Quality of research	Quality of technological infrastructure		SERVICES
7	Organization	Organization		
8	Sustainability of the institution	Sustainability of the institution		
9	"Reputation" /impact	"Reputation" /impact		

CRITERION	INDICATORS (from project partners)	Importance (Delphi M)	INDICATORS (from Delphi & EMEM review)
Quality of learning experience	Quality of pedagogy /methodology	4.25	Quality of interactions between educators and learners
	Quality of course / learning design	4.22	Specific training courses for teachers accessible throughout the academic year
	Quality of assessment	4.14	Quality of online tutoring/mentoring/assistance
	Quality of learning materials	4.12	Tracking of online interactions
			Quality of interactive environments and collaborative environments
Quality of research	Use of research for improvement and innovation (research-based projects…)	3.54	Involvement of students in research activities as co-partners
	Research output (in terms of publications, visiting scholars, ….)	3.30	Percentage of staff that is active in research
	Research in online teaching & learning (research groups, research projects, etc.)	3.30	Interdisciplinary impact
	Teaching staff engaged in research in online education	3.23	Effective application of research findings and best practices in teaching
	Internal centres devoted to research in online education	3.17	
	Research in other disciplines	2.90	

CRITERION	INDICATORS (from project partners)	Importance (Delphi M)	INDICATORS (from Delphi & EMEM review)
Organization	Credit transfer system aligned with national (and /or European) systems and operates bidirectionally	4.00	Ability of managing time and avoiding workload
	Examination policies able to cater with for the needs of e-learning courses	3.55	Teachers' collaboration with external partners, experts, institutions to deploy innovative and authentic learning settings
	Existence of a complaints and appeals system for learners	3.18	Teachers' interdisciplinary and technical collaboration in forms of co-teaching and co-working
	Non instructional support services (providing assistance for admission, financial issues, registration, enrollment,etc.)	3.05	Student satisfaction
	Structures such as libraries, labs, etc.	3.02	
	Decentralized structures on the territory	2.37	
Quality of teaching	Profile of the teaching staff	3.46	
	Teacher assessment and quality control	3.40	
	Standards for regulating teacher-student interactions (teachers' response time, modalities, timeliness of feedback provision for tests, etc.)	3.40	
	Involvement of academics in teaching	3.23	
Sustainability of the institution	Institutional strategic plan for online education	4.04	Clear indications about sustainability issues when evaluation is concerned
	Overall coherence of program design and provision (interconnections among courses, flexibility of the design, clarity of program design, …)	3.56	Pedagogical training of teacher and training concerning the code of ethics
	Resources (including financial ones) specifically devoted to the online program	3.49	Connections with other institutions at local level
	Sustainability of the portfolio of programmes	3.26	Interconnections with global institutions
	Existence of standardized workflows for program /course /material design and development	3.08	Funding
	Clear policy regarding OERs and MOOCs	3.32	
	Size of the institution	2.18	
"Reputation" /impact	Job opportunities for graduates	3.35	Placement in international and national rankings
	Internship and mobility opportunities	3.02	Relationship with the territory
	Social impact	3.08	Visibility on academic social networks
	Institutional image	3.38	Number of students
	Communication strategies	3.20	Attrition rate
	Organization of events, conferences, etc.	2.55	Number of students completing the courses
			Standard of assessment
			Representation on national forums
			Existing traditions of the institution
			Public profile of staff, students, and alumni
			Credibility of the institution

CRITERION	INDICATORS (from project partners)	Importance (Delphi M)	INDICATORS (from Delphi & EMEM review)
Quality of student support	Mentoring, tutoring activities to support student learning	3.74	Encouraging peer learning
	Technology support	3.64	Possibility of personalisation and self-regulation opportunities
	Orientation services to help learners taking decisions about their learning path	3.49	Dematerialization of processes and smart information system use (e.g. smartphone app)
	FAQ systems, Helpdesk, or sim. for learners	3.15	
	Regularity of information update	3.08	
	Support to student community building	2.92	
	Newsletters or sim. to allow communication from the institution to the learner	2.49	
	Support to alumni community building	2.31	
Quality of teacher support	Technology support	3.67	Recognition of teachers' efforts in their career (e.g. tenure, economic advantages)
	Opportunities for teaching staff to be trained in online education	3.59	Availability of good quality sample courses with increasing levels of teacher involvement
	Support to teachers and lecturers (for example provided via a specific department devoted to this)	3.44	Professional development systems to support professional learning on teaching and learning in higher education
	FAQ systems, Helpdesk or sim. for teachers	3.00	Updating of support materials
	Regularity of info update	2.95	
	Support to teacher community building	2.92	
	Newsletters or sim. to allow communication from the institution to the teaching staff	2.44	
Quality of technological infrastructure	Consistency/ robustness of the platform	3.73	Interoperability with external open sites (e.g., social media, DropBox, Google Drive)
	Capability to support evaluation	3.62	Technician training
	Capacity and concurrence of the platform (i.e. number of people that the infrastructure can host, even at the same time)	3.60	Compliance with international standards for interoperability between LMSs (Learning Management Systems), information and teaching/learning materials exchange (LTI, SCORM, …), and Single sign-on (SSO) access control
	Adequacy of the technical support (provided by department/ ad hoc team/or external company in charge of the technology)	3.60	Ability for learners and teachers to innovate and use newer add-ons/plugins
	Flexibility and scalability of the platform in view of future changes /new emerging needs	3.58	The cost of providing the infrastructure versus the cost of providing people to support student interaction with the infrastructure
	Adequacy of data security mechanisms	3.55	
	Compliance of interfaces with usability and accessibility standards	3.52	
	Capability to support interaction and teamwork	3.50	
	Existence of a plan for system maintenance and contingency management	3.50	
	Capability of the platform to satisfy different user needs (personalization capabilities, etc.)	3.45	

Pozzi et al., 2017

Appendix 4. Results from student participants

A4.1 Output

A number of 71 students took part to the first round of the Delphi Study (44 females, 25 males, 2 undisclosed; age 31.40 ± 11.05, range 18-65).

Fifty-five participants reported Italy as their country, while the rest of the sample reported the UK or Spain (7 each). Two participants did not disclose their home country.

A4.2 Output

Indicators' importance

The quartile distributions for the ratings of importance of the proposed indicators are reported in Table 2.

Table 2 – Indicators' importance

Criterion	Indicator	1st quartile	Median	3rd quartile
Quality of student support	Orientation services to help learners taking decisions about their learning path	3	4	4
	Mentoring, tutoring activities to support student learning	3	4	4
	Support to student community building	3	3	4
	Support to alumni community building	2	3	4
	FAQ systems, Helpdesk, or sim. for learners	3	3	4
	Newsletters or sim. to allow communication from the institution to the learner	3	3	4
	Regularity of information update	3	4	4
	Technology support	3	4	4
Quality of teacher support	Opportunities for teaching staff to be trained in online education	3	4	4
	Support to teachers and lecturers (for example provided via a specific department devoted to this)	3	3	4
	Support to teacher community building	2	3	4
	FAQ systems, Helpdesk or sim. for teachers	3	3	4
	Newsletters or sim. to allow communication from the institution to the teaching staff	2	3	3
	Regularity of info update	3	4	4
	Technology support	3	4	4

Quality of technology infrastructure	Compliance of interfaces with usability and accessibility standards	3	4	4
	Consistency/ robustness of the platform	3	4	4
	Capacity and concurrence of the platform (i.e. number of people that the infrastructure can host, even at the same time)	3	4	4
	Flexibility and scalability of the platform in view of future changes /new emerging needs	3	4	4
	Capability of the platform to satisfy different user needs (personalization capabilities, etc.)	3	3	4
	Capability to support interaction and teamwork	3	3	4
	Capability to support evaluation	3	4	4
	Adequacy of the technical support (provided by department/ ad hoc team/or external company in charge of the technology)	3	4	4
	Adequacy of data security mechanisms	3	4	4
	Existence of a plan for system maintenance and contingency management	3	4	4
Quality of the learning experience	Quality of course / learning design	4	4	4
	Quality of pedagogy /methodology	3	4	4
	Quality of assessment	3	4	4
	Quality of learning materials	4	4	4
Quality of research	Research output (in terms of publications, visiting scholars, ...)	3	3	4
	Research in online teaching & learning (research groups, research projects, etc.)	3	4	4
	Teaching staff engaged in research in online education	3	3	4
	Internal centres devoted to research in online education	3	3	4
	Research in other disciplines	3	3	4
	Use of research for improvement and innovation (research-based projects...)	3	3.5	4
Quality of teaching	Profile of the teaching staff	3	4	4
	Teacher assessment and quality control	3	4	4
	Involvement of academics in teaching	3	4	4
	Standards for regulating teacher-student interactions (teachers' response time,	3	4	4

	modalities, timeliness of feedback provision for tests, etc.)			
Quality of organization	Decentralized structures on the territory	3	3	4
	Structures such as libraries, labs, etc.	3	4	4
	Non instructional support services (providing assistance for admission, financial issues, registration, enrollment,etc.)	3	4	4
	Examination policies able to cater with for the needs of e-learning courses	3	4	4
	Credit transfer system aligned with national (and /or European) systems and operates bidirectionally	3	4	4
	Existence of a complaints and appeals system for learners	3	4	4
Sustainability of the institution	Institutional strategic plan for online education	3	4	4
	Size of the institution	2	3	3
	Overall coherence of program design and provision (interconnections among courses, flexibility of the design, clarity of program design, ...)	3	4	4
	Existence of standardized workflows for program /course /material design and development	3	3	4
	Sustainability of the portfolio of programmes	3	4	4
	Resources (including financial ones) specifically devoted to the online program	3	4	4
	Clear policy regarding OERs and MOOCs	3	3	4
Quality of reputation	Job opportunities for graduates	4	4	4
	Internship and mobility opportunities	3	4	4
	Social impact	3	3	4
	Institutional image	3	4	4
	Communication strategies	3	3.5	4
	Organization of events, conferences, etc.	2.5	3	4

The results from students confirm that most indicators are deemed very important for assessing their criterion. No indicator has a median below 3 (on a scale ranging from 0=not at all important to 4=extremely important).

Inter-rater agreement, measured using a modification of Fleiss' kappa for ordinal data (Marasini, Quattro, & Ripamonti, 2014) is of 0.55 (percentile-bootstrapped 95% confidence interval of [0.52, 0.58]), same as for the main sample.

Students, too, proposed additional indicators for each criterion.

The following list omits indicators we already considered for a different criterion (e.g. quality of teaching has been suggested as an indicator of reputation, but is already its own criterion).

- For Quality of student support, suggested indicators were: response time to queries; accuracy of responses; student satisfaction with responses.
- For Quality of teacher support, suggested indicators were: availability of student feedback; opportunity for remote staff to network at events to establish a community/reduce the sense of isolation; promotion of sharing of good practices; orientation programme.
- For Quality of technological infrastructure, proposed indicators included: response to system outages; ability to respond to changes in 3rd party items (e.g. Oss updates); cross-platform support; user friendliness; screen reader support; security from external access.
- For Quality of the learning experience, new suggestions included: views of external examiners and, where appropriate, accreditation bodies; availability of learning material in different formats; personalizability of methods of delivery of learning material; coherence between learning material and learning goals.
- For Quality of research, new propositions were: relevancy of the research to the curriculum; engagement of students in research.
- For Quality of organization, participants only suggested to consider exam organization (seats availability, times, staff present, etc.).
- For Sustainability of the institution, suggestions were: employability for graduates; maintenance of student number.
- Lastly, for Quality of reputation participants suggested to consider: students' opinion; maintaining academic standards of qualification.

Relative importance of criteria

The criteria rankings were analyzed using Thurstone Case V scaling (Thurstone, 1927; see Figure A1).

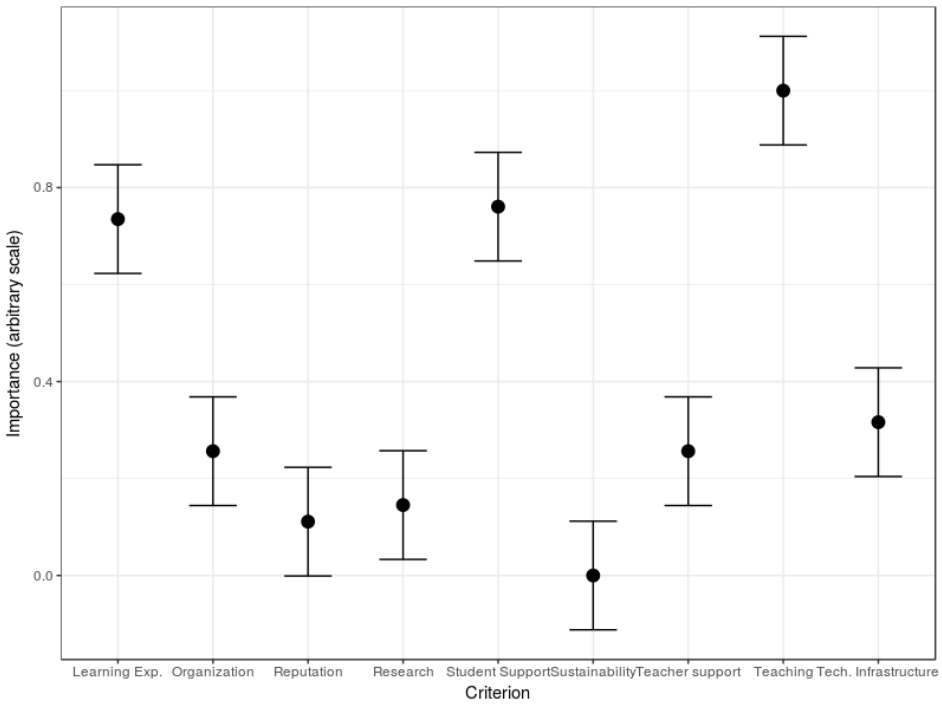


Fig.7 - Relative importance of the evaluation criteria

Looking at the obtained scaling, we can see that Quality of the learning experience is considered the most important criterion.

Quality of teaching and Quality of student support are considered of medium importance; the other criteria form the lowest tier of indicators of quality.

The ranking obtained is different from the one obtained from the main sample. In particular (and unsurprisingly), Quality of student support is ranked considerably higher by students. Additionally, the lowest tier encompasses more criteria than in the main sample, suggesting that students consider three criteria as very important (quality of the learning experience, of teaching, and of student support) while not clearly differentiating between the others.