Project Successful Deployment

A Method for Evaluating the Success of Digitalization Projects

https://doi.org/10.3991/ijoe.v18i14.35087

Giovanni Romagnoli(✉), Giovanni Esposito, Davide Reverberi
University of Parma, Parma, Italy
giovanni.romagnoli@unipr.it

Abstract—In this paper, we present a method for evaluating the success of digitalization projects, namely the Project Successful Deployment (PSD). With the term digitalization, we mean the use of digital technologies and digitized information to create value in new ways and to benefit from them. The existing methods for project evaluation emphasize the capability of a project to deliver its results by respecting times and costs. The method we propose, instead, suggests evaluating projects by means of its external dimensions, namely the functionalities and quality of the deliverables. These external project dimensions are reflected on the project scope, and thus evaluate the requirements of the deliverables, and the degree to which the deliverable meet their quality objectives The method is composed of a set of matrixes, and it uses a structured procedure to define and refine its items and their weights, by means of a panel of experts. It has been applied to a practical case study, a digitalization project of a network of research and teaching laboratories. The method allowed a structured project evaluation, and the practical case study showed strengths and weaknesses of the PSD model, which proved to be robust and effective, in providing a timely evaluation of the project.

Keywords—project management, project scope, digitalization, functionality, quality criteria, evaluation methods.

1 Introduction

The covid-19 pandemic has given a significant acceleration to the need of providing remote access to the facilities of both public and private organizations in various service and industry sectors, to guarantee continuity of operations and of the delivery of products and services [1]. To contain the spread of the new coronavirus, in fact, several countries have adopted preventive measures to limit social interactions as much as possible [2]. As such, these measures have produced contrasting impacts on different activities. Whereas the online shifting of several activities has been smoother and less remarkable (such as office work and university classes, just to provide two examples), other kind of activities, requiring physical presence or the direct interaction with physical resources, have experienced a much greater challenge (see for example shop floors, warehouses, plants, and laboratories).

In the last decades, a great effort has been spent in researching issues related to providing remote access to physical laboratories, as several studies report (see for example

[3], [4]. These type of labs are often labelled with the general term of non-traditional labs (NTLs, [5]), an umbrella term that encompasses several different types of labs, such as online, remote, virtual, and hybrid. This significant amount of studies, which are mostly related to Engineering education and research [6], can be partitioned in two main research lines [7]: the former focuses on the educational aspects of these labs, aiming to validate their didactical proposition, and the latter deals with the design and implementation issues related to NTLs, such as the network architecture, equipment automation, safety and security of people, assets, and data.

With respect to the latter research line, we note that NTLs are frequently delivered by medium-to-long term digitalization projects that provide remote or hybrid access to pre-existing hands-on labs or develops brand new virtual or online labs [8], [9]. In the following, we will refer to the term 'digitalization' as in [10], that is the use of digital technologies and digitized information to create value in new ways and to benefit from them. In a recent review of NTLs and lab network initiatives, [5] noted that: (i) NTLs have been very prolific in the last decades; (ii) NTLs and lab networks have mostly been funded by public bodies, whose fundings almost reach 70% of their results; (iii) publicly funded labs, however, experienced much shorter durations over time, with an average duration of these type of initiatives of 6.4 years.

Given the importance and the research attention on this topic, as well as the significant percentage of digitalization projects that are publicly funded, it is quite surprising to note that only very few studies evaluate the success of such digitalization projects from different standpoints. For instance, the work of [11] introduces the (i) 'cost' driver for evaluating the implementation success, and (ii) the diffusion of digital labs through the Rogers' Diffusion of Innovation Model. But mostly, the works available in the literature on this topic deal with the issue of evaluating the reliability of NTLs, namely digital online labs, either from the educational point of view or from the effectiveness of the technical implementations. In fact, as emerges from [12], research on the topic revealed two aspects. First, since 2000, one of the main concerns of digital online labs is the effectiveness in students' satisfaction and knowledge gain. Second, from 2015 the focus widened into practical implementation of remote laboratories to provide the diverse learners with e-learning environment, requiring standardized practices to integrate platforms to practical scenarios. This has been mainly possible because of the advent of the 4th industrial revolution, the so-called Industry 4.0 and related disruptive technologies [13].

At the cutting edge of Industry 4.0, digitalization has become a recurring goal, and sometimes even a buzzword, not only in education but in everyday life, and especially in manufacturing: it is, in fact, one of the biggest and most trendy challenges of manufacturing and services [14]. It is generally agreed, however, that several companies find this digital transition quite challenging, and that this topic creates concern to many a manager. One of the possible causes behind this fact is the lack of standardized instruments for following and handling this digital transition [15]. Indeed, several studies have discussed the topic, and interesting results have been achieved for setting drivers and barriers to this transition [16]–[18]. We note, however, that these studies are focused on the identification of maturity models or framework for addressing the digital transition, rather than aiming at providing a practical method for enabling and supporting the evaluation of digitalization.

Therefore, methods for evaluating completeness and quality of digitalization projects remain a rather unexplored field. From our point of view, the problem relates to the evaluation of project success, and hence it can be approached as a Project Management (PM) problem. Reference [19] define a project as '*an endeavor in which human, material and financial resources are organized in a novel way, to undertake a unique scope of work, of given specification, within constraints of cost and time, so as to achieve beneficial change defined by quantitative and qualitative objectives*'. PM can be defined as the use of specific knowledge, skills, tools and techniques to deliver something of value to people [20].

Therefore, the present paper aims to answer to the following research questions:

(RQ1) – Is it possible to devise a method for evaluating the success of digitalization projects? If so, which model could be used for this goal?

In particular, the present paper aims at designing a method for evaluating the degree of completeness and the level of success achieved by a digitalization project, namely the Project successful Deployment (PSD).

(RQ2) – How should this method work, and which specific details must be considered to devise it?

Namely, the present paper proposes a method to identify items that might prove if the digitalization project delivered its scope, and the degree to which project benefits have been achieved, thus delivering value to stakeholders.

We note that the method and the model that we propose are derived from and validated in a specific environment, namely a digitalization project of a network of research and teaching laboratories. Still, the approach reported in this study, the structure of the model, as well as the method for identifying specific details can be generally used in projects evaluation, and especially in the evaluation of digitalization projects.

The reminder of the paper is organized as follows. The theory and instruments of PM that are useful for this study are briefly introduced in Section 2. Section 3 provides the model of the PSD. We introduce the case study where the model is applied in Section 4, alongside with the specific details of the application, and the results from the adoption of the PSD to the case study. Finally, Section 5 addresses conclusions and outlines possible future directions of research.

2 A review of evaluation models for project success

In this section, we provide a review of the literature on existing structures for evaluating project success. More precisely, we aim at understanding how these structures might be designed and used to assess and evaluate projects and their results, with a specific focus on digitalization projects.

Reference [21] claim that project success evaluation models are not suitable for all project types, and moreover the project success measurement system does not usually fit with systems used by project individuals, mainly the project manager. The authors propose their model in three distinct project success dimensions, namely (i) the project manager performance in achieving the project plan, (ii) the project owner performance in realizing the business case, and finally (iii) the investment performance of the project for its funder. The model is admittedly theoretical.

Reference [22] propose the two-stage Construction Project Productivity evaluation framework to indicate site efficiency and utilization effectiveness, and then taking into considerations the productivity of both the construction and post-construction stages. The framework is qualitative and focused on Hong Kong construction industry.

Reference [23] propose machine-learning based method for monitoring and controlling the development process at different stages of the life cycle of software development using Agile approach.

One of the main approaches for evaluating project success is that of adopting Multi-Criteria Decision-Making method for the assessment: [24] use M-TOPSIS for evaluating the success of construction projects, according to suitable identified success criteria. Authors admits that the method is only suitable when massive data from the project can be analyzed and considered. Reference [25] use ANP for proposing a framework that provides project stakeholders with a forecasting and diagnostic tool to evaluate progressively and objectively the project chances of success to assist in improving overall project performance. Reference [26] use an Evolutionary Fuzzy Hybrid Neural Network for monitoring project cash flow.

An interesting approach is provided by [27], whose aim is to develop an ex-post evaluation procedure for Public-Private Partnership projects. The authors identify 5 sources of complexities to consider, namely (i) large size and technically complex projects, (ii) multiple perceptions of the impacts, (iii) vague and uncertain understanding of 'public interest', (iv) long time horizon for the evaluation, and (v) political and ideological drivers that are relevant and difficult to address. However, we note that this study cannot be adapted to our case study, due to several missing data.

Consequently, it is opinion of the authors that present studies, although noteworthy, do not provide a comprehensive picture to support the evaluation of digitalization project. Firstly, one of the main drivers considered to this aim, and often the only one, is the 'project cost'. Also, another concern refers to the use of indicators and tools for evaluating (i) whether milestones are on time, and (ii) the adherence of the project progress with its baseline. Therefore, the focus on the results of the project, namely the effectiveness and the quality of its deliverables, seems to be missing. Finally, the evaluation models we describe above require timely data on project progress, and they are often performed by the project sponsor, or by the project performing organization. Thus, the reliability of these evaluations can be undermined if the project cannot produce enough data on time, and these models often miss the point of view of users and other stakeholders.

3 The project success deployment PSD

3.1 The PM approach for the PSD

We decided to follow a different approach, namely identifying the dimensions of the success of a project. Reference [28] states that the project is considered an overall success if it meets the technical performance specifications, and if key people of the project team and related stakeholders get a high level of satisfaction concerning the project deliverables. Reference [29] categorizes these into two dimensions. The 'external' dimension relates to the adherence of characteristics of deliverables with the mission to be performed and is translated into functionalities and quality of the deliverables. The 'internal' dimension relates to the efficiency of project processes and is translated into three drivers of cost – the budget adherence, time – the respect of schedules, and scope – the objective to achieve. The external dimension is measured after the project closure and release, while the internal dimension must be measured during project execution. According to [30], a significant amount of literature approaches the problem of monitoring the internal dimensions of projects, whereas very few studies discuss methods for evaluating the external dimension, composed of 'Functionalities' and 'Quality' of deliverables. They approach the problem by encompassing external dimensions within the internal dimension of the scope. As a results, internal dimension of 'Functionality' translates into the requirements of the deliverable defined by the project team, and the 'Quality' translates into the characteristics of the project deliverable expected or required by the customer, as it is shown in Figure 1. The authors of [29], however, do not provide guidance for reflecting the external dimension to the scope, as well as no method to evaluate project success.

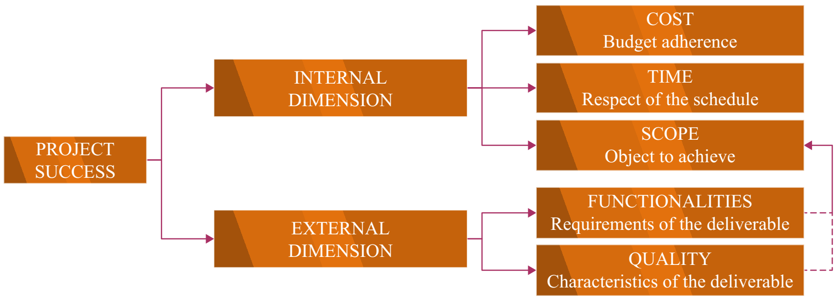


Fig. 1. Translation of external dimension into the internal dimension of the scope, according to [29]

We move from this gap, and from the perception that several tools and indicators do exist for monitoring the project in terms of costs and time, as a wide set of project management books can confirm. On the contrary, we experienced a lack of similar instruments for the evaluation of the project scope.

3.2 The structure of the PSD

In this Section, we describe the structure of the PSD and the quantitative framework for computing the result of the project in terms of success in delivering its scope. To this aim, we start from the following definitions, which will be used in the reminder of the paper:

- Functionality – an action, operation, capability, or usefulness that a project aims to deliver from its proposal phase, and thus that one or more of its deliverables are expected to fulfill. We used the term 'Aggregated Functionality' to group more than one Functionality at a lower level of detail
- Quality Criteria – the specific characteristics or aspects that will be selected, tested, and measured to confirm that the quality objectives of the functionality have been met. Also, we used the term 'Aggregated Quality Criteria' to group a set of Quality Criteria at a lower level of detail
- Method – a computational tool for evaluating the degree to which deliverables (or the whole project) and their Functionalities meet their quality objectives
- Items – elements to be considered by the method for evaluating the project. This category comprehends both Functionalities and Quality Criteria
- Ratings – numerical values expressing a qualitative judgement
- Indicators – Results of the computations of ratings expressed in a useful manner
- Model – Method filled in with items for computing identified indicators.

Also, the list of symbols and quantities used in the paper is reported in Table 1. The Project Success Deployment is inspired to the Manufacturing Cost Deployment of [30], as described by Braglia et al., who adapted this tool to develop their Project Cost Deployment [31]. We note that the approach of these two methods is similar, as it consists of different matrixes, who analyze and further details the results achieved at the previous step (i.e., taken from the previous matrix). While the reader is referred to the previously mentioned papers for the details of these methods, we report below the structure of the PSD, which can be decomposed in 4 different matrixes:

- A-Matrix relates the deliverables (or some level of the Work Breakdown Structure – WBS) to the aggregated functionalities that those deliverables are planned to achieve. As such, this matrix details the 'Aggregated Functionalities' (columns) that the project aims to achieve in its different scope areas, i.e., 'Deliverables (or WBS branches)' (rows). If an Aggregated Functionality expresses one of the goals of a deliverable, the related cell is ticked off (e.g., by means of a green-colored background), as this action makes it simpler to arrange the next matrix.
- B-Matrix specifies the Aggregated Functionalities in suitable Functionalities. The matrix is arranged with the Deliverables (or WBS branches) and the related Aggregated Functionalities on the rows. We note that green- and red-colored background colors allow a simple listing of the results of the A-Matrix on the rows of the B-Matrix. Also, we stress the fact that the same Aggregated Functionality can be listed in more than one Deliverable. On the columns, the B-Matrix lists the Functionalities, that is a more detailed level of what each deliverable is aimed at, with respect to a given Aggregated Functionality. Again, the B-Matrix uses red- and green-colored background colors in its cells to link the Aggregated Functionalities to the specific Functionalities as it is reported in the next step.
- C-Matrix stresses the Aggregated Quality Criteria that shall be used to evaluate the functionalities obtained from the B-Matrix. At this step, in fact, 'broad' Quality Criteria are connected to the Functionalities of each Deliverable. Thus, this matrix transposes the Functionalities on the rows, connected to the respective Aggregated Functionalities. Again, we remind that one Functionality can be listed in more than one Aggregated Functionality. The matrix displays the Aggregated Quality Criteria in its columns. If an Aggregated Quality Criteria expresses the characteristics of the Functionality, the related box is ticked off.
- Finally, the D-Matrix relates the specific Quality Criteria to the Functionalities. As such, it supports the evaluation of how much each Functionality meets the Quality Criteria. The D-Matrix reports the detailed Functionalities and detailed Quality Criteria in its rows and columns, respectively. Here, again, each Quality Criterion can be connected to more than one Aggregated Quality Criterion. We note that the relationship between Functionalities and specific Quality Criteria (belonging to an Aggregated Quality Criteria) is simply traceable by means of symbols inserted in C-Matrix.

The schematic representation of the PSD is reported in Figure 2. Green and red cells refer to the link between WBS branches and Aggregated Functionalities, as well as these and single Functionalities. If a relationship exists, then the cell is green colored. Similarly, if a Functionality is evaluated with respect to a Quality Criterion, and then detailed by relative Quality Criteria, the correspondent cells are filled in with a X. The reader can note that, although a correspondence Functionality-Aggregated Quality

Criteria, it is not taken for granted that a single Quality Criteria does relate to a Functionality even if it belongs to the related Aggregated Quality Criteria. The reader can see this in Figure 2 where, just as an example, this happens to Functionalities 'F1' and 'Fn' in the Aggregated Functionalities 'AF1' and 'AFn'.

Once obtained the D-Matrix, with Functionalities and related Quality Criteria, it is possible to proceed as it is described in section 3.3.

Table 1. List of symbols and quantities used in this paper

Quantity	Description
$N = N_f + N_q$	Initial amount of identified items of the PSD (N_f Functionalities and N_q Quality Criteria)
$M = M_f + M_q \leq N$	Final number of selected items of the PSD (M_f Functionalities and M_q Quality Criteria)
$i = \{1, 2, \ldots, N_f\}$	Functionality at the beginning of the selection process. At the end it is $i = \{1, 2, \ldots, M_f\}$
$h = \{1, 2, \ldots\}$	Aggregated Functionalities. It does not matter to computations how much they are
H	Number of elements in the Aggregated Functionality h at the beginning of the selection process. At the end it is H_f
$j = \{1, 2, \ldots, N_q\}$	Quality Criteria at the beginning of the selection process. At the end it is $j = \{1, 2, \ldots, M_q\}$
$t = \{1, 2, \ldots\}$	Aggregated Quality Criteria. It does not matter to computations how much they are
T	Number of elements in the Aggregated Quality Criterion t at the beginning of the selection process. At the end it is T_q
$\begin{cases} k = i \forall i \\ k = j \forall j \end{cases}$	Variable substitution to simplify computation description. If items are Functionalities, $k = i$. If items are Quality Criteria, $k = j$
$\iota = \{1, 2, \ldots, \mathrm{I}\}$	Experts involved in the framework for selecting the items of the PSD
$r_{\iota\kappa} = \begin{cases} r_i \\ r_j \end{cases}$	Rating of importance of the *kth* items (*ith* Functionality or *jth* Quality Criterion) expressed by the *ιth* expert
$\mu_\kappa = \frac{\Sigma_{\iota=1}^{\mathrm{I}} r_{\iota\kappa}}{\mathrm{I}} = \begin{cases} \frac{\Sigma_{\iota=1}^{\mathrm{I}} r_{\iota i}}{\mathrm{I}} = \mu_i \\ \frac{\Sigma_{\iota=1}^{\mathrm{I}} r_{\iota j}}{\mathrm{I}} = \mu_j \end{cases}$	Indicator of the importance expressed by the I experts
$K \leq M$	Number of items belonging to a single WBS branch (or to a single deliverable)
$\bar{\mu} = \frac{\Sigma_{\kappa=1}^{\mathrm{K}} \mu_\kappa}{\mathrm{K}}$	Indicator of the average value of means of ratings belonging to the same K

(Continued)

Table 1. List of symbols and quantities used in this paper *(Continued)*

Quantity	Description
k_C	Cohen's kappa coefficient
$W_{ij} = \mu_i * \mu_j$	Weight of importance of the Functionality *i* and the level of quality *j* required by the project
a_{ij}	Single rating of judgement expressed by each *ιth* expert
I*	Number of experts involved in the evaluation of the project success
$a^*_{ij} = \begin{cases} a_{ij} \\ \overline{a_{ij}} = \Sigma_{\iota=1}^{I^*} \frac{a_{ik\iota}}{I^*} \end{cases}$	Degree to which the Functionality *i* meets the Quality Criterion *j*. It values a_{ij} if expressed as collective judgment, $\overline{a_{ij}}$ if each expert expresses its own rating
$p_{ij} = w_{ij} * a^*_{ij}$	Indicator of the result achieved by the project with respect to the *ith* Functionality and the *jth* Quality Criterion
$b_{ij} = w_{ij} * 5$	Benchmark to compare the best implementation possible of the *ith* Functionality having the characteristic of the *jth* Quality Criterion
$S_{ht} = mean(p_{ij})_{ht} = \frac{\Sigma_{h=1}^{H_f}\Sigma_{t=1}^{T_q} p_{ij}}{H_f * T_q}$	Mean value of results p_{ij} of the *hth* Aggregated Functionality and the relative *tth* Aggregated Quality Criterion
$S_{ht}^{bench} = mean(b_{ij})_{ht} = \frac{\Sigma_{h=1}^{H_f}\Sigma_{t=1}^{T_q} b_{ij}}{H_f * T_q}$	Mean value of benchmarks b_{ij} of the *hth* Aggregated Functionality and the relative *tth* Aggregated Quality Criterion

3.3 Identification of items and their weights

The identification of a cluster of N items, that is Functionalities and Quality Criteria, each of which can be related to specific areas of the project, in terms of deliverables or branches of its WBS, can only be performed by a designated project team, according to the project organization, the specific field of expertise and its expected deliverables. The selection of $M \leq N$ items to fill in the PSD structure, however, as well as its validation, requires a rigorous quantitative process.

Identification of the panel. The Cochran's formula for small sample size is used for identifying the sample size of I experts in the target population, according to the desired level of confidence [32], [33]. It is important to select experts from different fields of specialization, to have different points of view about the project success.

1st Delphi Round – Ratings of the items' suitability, and robustness of the rating scale. Each ι expert is called to express ratings of suitability $S_{\iota\kappa}$ for each item κ, via a first round of Delphi method. The round can be performed in two steps, one for the Functionalities and one for the Quality Criteria, or in just one step for both. Each expert expresses its rating individually and independently. The rating scale can be adopted arbitrarily. However, Likert scale is more suitable, since its validation is performed by means of the Cronbach's alpha α as in [34] – once that all ratings are collected. Process shall be repeated for both $\kappa = i$ Functionalities and $\kappa = j$ Quality Criteria, respectively.

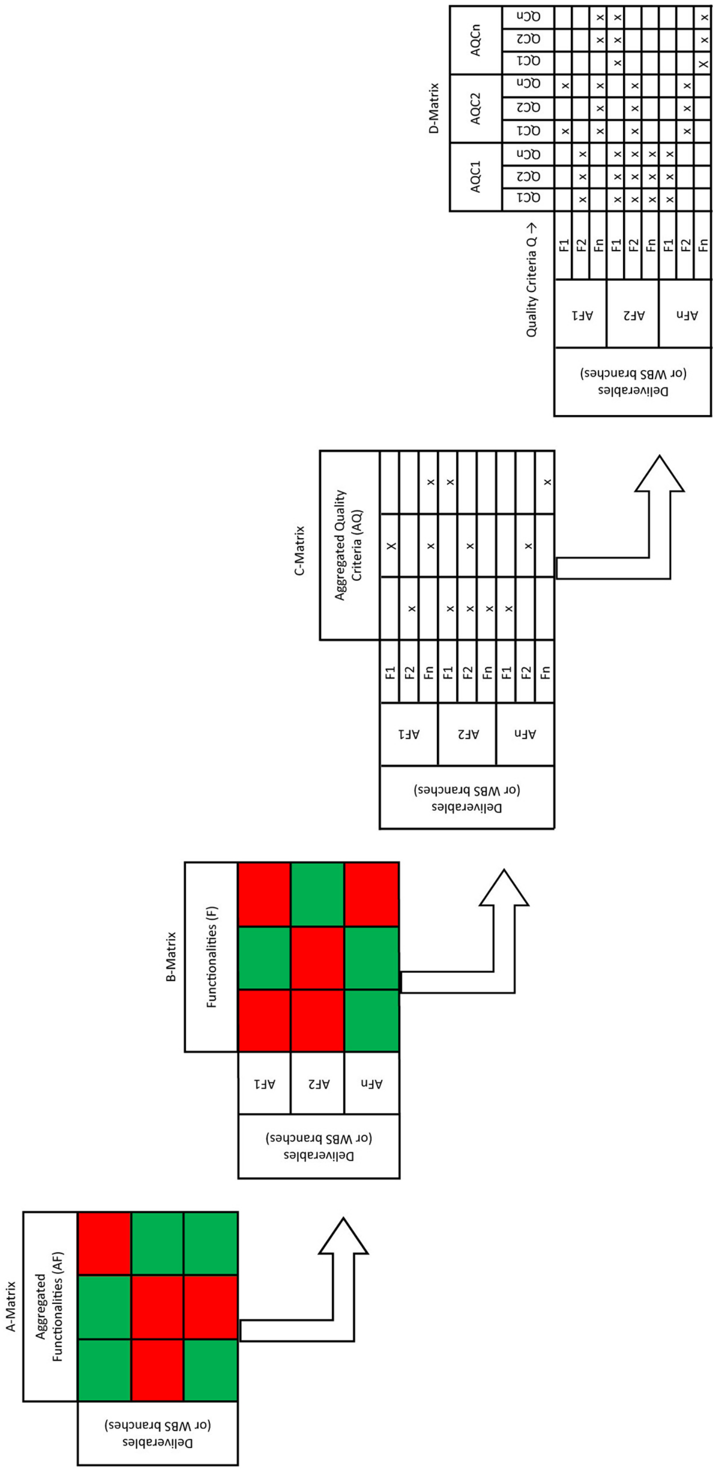


Fig. 2. Schematic representation of the PSD

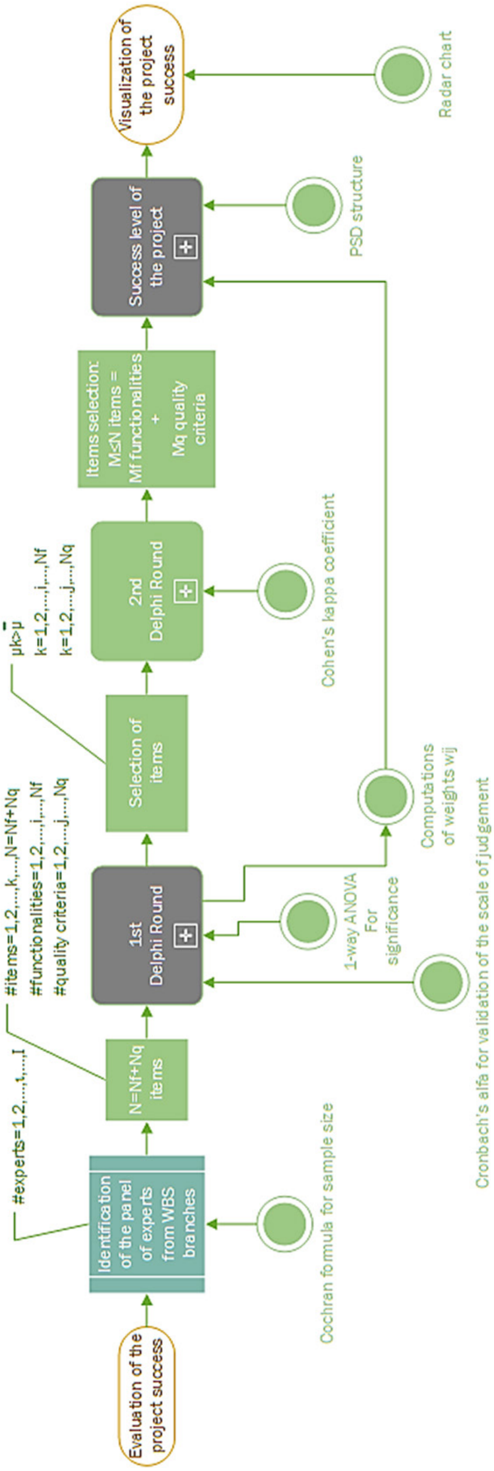


Fig. 3. Overall framework and sub-processes (in light-green boxes) for identification of items in the PSD

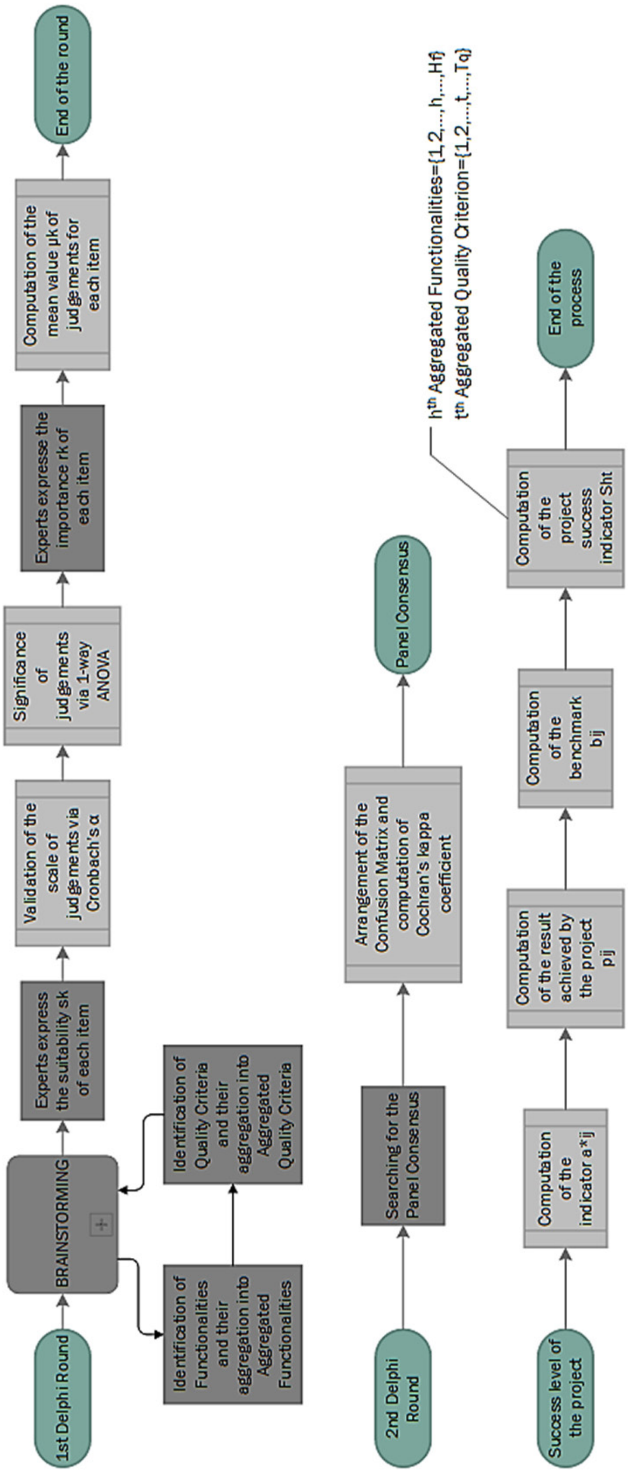


Fig. 4. Detail of main Delphi processes for panel consensus

Significance of ratings with experts' background. Two one-way ANOVA tests must be employed to compare the views of project experts, expressed by their ratings, with different backgrounds. The former test computes the significance of the correlation between the experts' field of expertise, and the specific item. The latter computes the significance of the correlation between single expert, and the aggregated items. With respect to this point, we note that some different levels of aggregation could be used (e.g., WBS branch, rather than Aggregated Functionalities). We leave the specific decision on which one to use to the project evaluators. Also, we chose here to perform two one-way ANOVA, instead of a two-way ANOVA because the independency of the two variables 'field of expertise' and 'aggregated items' cannot be easily proved.

Selection of items. Each expert ι of the panel I is then called to rate the importance of each item κ for the project success $r_{\iota\kappa} = \begin{cases} r_i \\ r_j \end{cases}$ depending on if the expert is evaluating importance of Functionalities or Quality Criteria, of course. Thus, the mean value $\mu_\kappa = \frac{\sum_{\iota=1}^{I} r_{\iota\kappa}}{I}$ and then the total means $\bar{\mu} = \frac{\sum_{\kappa=1}^{\kappa=K} \mu_\kappa}{K}$ are computed, separately for each ambit holding K items. If $\mu_\kappa \geq \bar{\mu}$ the item is held, otherwise is discarded (see also [34], for a similar approach). In this case, κ refers to both $\kappa = i$ Functionalities and $\kappa = j$ Quality Criteria, respectively, and the process is repeated in two independent instances. It is noted that the same ratings expressed in the 1st Delphi round can be used, or experts can express a new judgment in a new scale, for instance for better detailing the experts' opinion. Searching for the panel consensus in the next phase will secure the validity of the ratings also in the case in which it is adopted a new scale of judgment.

2nd Delphi Round – Panel Consensus on item selection. A second round of Delphi is then performed with the same I experts, searching for the panel consensus on item selection. Each expert ι expresses his agreement with confirmation and elimination of items with a 1 and 0 input, and then the Cohen's kappa coefficient k_C is computed from the confusion matrix, and the threshold for good agreement level is set to $k_C \geq 0.60$ (see also [35]). We used C subscript for avoiding confusion with the k index previously introduced.

Computations of weights w_{ij} and results achieved by the project. The product of the mean values of ratings of the *ith* Functionality and the *jth* Quality Criterion, namely $\mu_{\kappa=i}$ and $\mu_{\kappa=j}$, is the weight $w_{ij} = \mu_i * \mu_j$ in the D-Matrix, and pre-multiplies the level of achievement of the project a_{ij}^* rated by the project team with respect to the same precisely *ith* Functionality and the *jth* Quality Criterion. Level of achievement a_{ij}^* can be expressed in diverse manner, as mean value $\overline{a_{ij}}$ of individual judgment of experts of the panel involved so far, as well as judgements of experts from another panel $a_{ij}^* = \overline{a_{ij}}$, or as a collective single judgment $a_{ij}^* = a_{ij}$. Judgments are expressed in a low-medium-high 1-3-5 scale. What the line pursued for judging a_{ij}^*, the product $p_{ij} = w_{ij} * a_{ij}^*$ is the indicator that represents the result achieved by the project with respect to the *ith* Functionality and the *jth* Quality Criterion. This result can be compared to the benchmark $b_{ij} = w_{ij} * 5$ which represent the best implementation possible of the *ith* Functionality having the characteristic of the *jth* Quality Criterion.

Visualization of project success. Visualization of the global result of the project is provided by means of a radar chart. For plotting data, it is computed the project success indicator as mean value of each result p_{ij} of the *hth* Aggregated Functionality, having and the relative *tth* Aggregated Quality Criterion, $S_{ht} = mean(p_{ij})_{ht} = \frac{\sum_{h=1}^{H^*}\sum_{t=1}^{T^*} p_{ij}}{H^* * T^*}$. Values obtained for each Aggregated Functionality with respect to the relative Aggregated Quality Criteria are plotted on the radar chart and can be compared to the scope supposed to deal with $S_{ht}^{bench} = mean\left(b_{ij}\right)_{ht}$.

Figures 3 and 4 summarize the framework detailed so far. We adopted a traditional flowchart convention, where light green boxes represent processes, while dark green ones are computations. Circles entering or exiting the boxes mean quantitative values for computation, the '+' sign in gray containers means sub-processes, and callouts are used for notes.

4 The DigiLab4U use case for the PSD application

DigiLab4U is a cross-Institutional network of IoT and Industry 4.0 lab infrastructures. The consortium, whose details can be found at the project website (http://digilab4u.com/), counts 5 founding institutions, and 9 more worldwide partners joined the consortium in 2021. The network was funded by the German Federal Ministry of Education and Research (BMBF) for developing the project 'Open Digital Lab for You', with the goal of creating an integrated and hybrid learning and research environment providing different types of labs for a digital offering reaching different kinds of users. The WBS of the project has been organized into three branches, which will also be labelled as pillars in the remainder of the paper. Identified pillars are arranged as follows:

1. Organizational: it investigates administrative, organizational, and commercial aspects of the project, such as trust, partners relationships and the potentiality for financial sustainability.
2. Didactical: it explores educational aspects, such as didactical methods and scenarios.
3. Technical: it investigates the several different technical aspects of the project, such as technologies selection and implementation, network architecture, and specific lab solutions.

The project team is composed of 23 people with different competences, that can be associated to the three WBS branches reported above.

Identification of the panel. Cochran's formula is applied. Since the population of experts is small, the sample size is calculated by the following formula:

$$n = \frac{n_0}{1 + \frac{\left(n_0 - 1\right)}{N}} = \frac{23}{1 + \frac{\left(23 - 1\right)}{23}} = 11$$

Where N is the total population of 23 experts, n_0 is the sample size obtained by the Cochran's formula. In this case it cannot be larger than same 23 experts and reversing the Cochran's formula this estimate fixes the confidence level at 92.5%, with probability of having positive answer from the experts $p_{\%} = 0.5$. In fact, we do not have much information on the subject to begin with, so we're going to assume that half of the panel agree with items selection. Area of expertise of the panel is transversal. Three experts have technical and didactical skills. One expert has technical and organizational skills. One has just technical skills. Two have just didactical skills. And finally, remaining 4 experts have just organizational skills. The skill matrix is then balanced by 5 experts for area of expertise.

1st Delphi Round. A first synchronous Delphi round was performed. Functionalities and Quality Criteria are introduced and then grouped into Aggregated Functionalities and Aggregated Quality Criteria, respectively, during a brainstorming session. For simplicity, we limited each item to a possible relation to a maximum of 2 different Aggregated items. The Technical pillar counts 2 Aggregated Functionalities and 13 Functionalities, as well as 8 Aggregated Quality Criteria and 32 Quality Criteria. The Organizational Pillar counts 3 Aggregated Functionalities and 7 Functionalities, 4 Aggregated Quality Criteria and 6 Quality Criteria. The Didactical Pillar counts 2 Aggregated Functionalities and 6 Functionalities, with 6 Aggregated Quality Criteria and 11 Quality Criteria. The full list of Functionalities and Quality Criteria is detailed in the Appendix (Tables A1–A6). After, the panel of experts was called to express their judgment on the suitability of items for the PSD. The experts expressed their judgements by means of a Likert scale, and Cronbach's α was calculated for all judgements, with all resulting values $\alpha > 0.60$, validating the results. With respect to Functionalities: $\alpha_{TF} = 0.80$, $\alpha_{OF} > 0.82$, $\alpha_{DF} > 0.69$ for the technical, organizational, and didactical pillar, respectively. With respect to Quality Criteria: $\alpha_{TQC} = 0.85$, $\alpha_{OQC} > 0.66$, $\alpha_{DQC} > 0.83$, respectively.

Significance of ratings with experts' background. The ANOVA tests were then performed for the significance of results of the 1st Delphi round ($\alpha = 0.05$). Experts were grouped according to their field of expertise, and the same applies to items, which were grouped according to the project pillar they belong to. Afterwards, the mean value of ratings for each item is computed according to the experts' background. Therefore, we computed the Mean value of Technical Experts' ratings, Mean value of Organizational Experts' ratings, and the Mean value of Didactical Experts' ratings for Functionalities (MTE_F, MOE_F, and MDE_F, respectively). The same applies to Quality Criteria, and thus MTE_QC, MOE_QC, and MDE_QC are computed. Similarly, we calculated the mean value of ratings given by experts per each pillar. For the Functionalities, these are the Mean value of Technical Functionalities (MTF), the Mean value of Organizational Functionalities (MOF), and the Mean value of Didactical Functionalities (MDF). For the Quality Criteria, these are MTQC, MOQC, and MDQC. Two one-way ANOVA tests were then conducted, to examine the effect of the experts' field of expertise on the evaluations provided per project pillar the items belong to. No statistically significant interaction was noted by the ANOVA analysis.

Selection of items. Items are then evaluated based on their importance for the project, and experts are called to express the importance of each item from their point of view $r_{\iota\kappa}$. Therefore, items are selected according to the rule that if the mean value of ratings of importance for the *kth* item μ_κ is equal to or higher than the mean value of all ratings of items belonging to the same project pillar $\bar{\mu}$, namely $\mu_\kappa \geq \bar{\mu}$, then the item is kept; otherwise, the item is discarded. We decided to adopt a rounded centesimal scale here (from 0 to 1), for better detailing experts' opinion and computing the weights w_{ij}

of next phases as a percentage. The mean value of importance of functionalities is equal to 0.69, 0.68, and 0.64 (for the technical, organizational, and didactical pillar, respectively), whereas the mean value of importance of quality criteria is equal to 0.65, 0.71, and 0.72 (for the technical, organizational, and didactical pillar, respectively). This operation resulted in 7 Functionalities and 2 Aggregated Functionalities for the Technical Pillar (i.e., LAB NETWORK and LAB HARDWARE), together with 14 Quality Criteria connected to 6 Aggregated Quality Criteria (i.e., SOFTWARE, TECHNIQUE, PROTOCOL, MIDDLEWARE, CUSTOMER-ORIENTED PROCEDURES, and SYSTEM EFFICIENCY). Also, the Organizational Pillar counts 3 Functionalities grouped into 2 Aggregated Functionalities (i.e., RESILIENCE and USE); 4 Quality Criteria clustered in 3 Aggregated Quality Criteria (i.e., USERS' INTENTION TO USE, USERS' ACCEPTANCE, SOLUTION & VIABILITY). Finally, the Didactical Pillar counts 3 Functionalities and 2 Aggregated Functionalities (i.e., LEARNING TOOL and LEARNING METHOD); as well as 6 Quality Criteria grouped into 5 Aggregated Quality Criteria (i.e., KNOWLEDGE, USABILITY, DIDACTICAL METHODS, COMMUNICATION, and FEEDBACK). We note that some Quality Criteria, albeit showing $\mu_\kappa \geq \bar{\mu}$, have not been considered due to the fact that the related Functionalities were discarded.

2nd Delphi Round. In the next phase, the panel consensus on items selected and discarded is searched. Experts are called to express their consensus on each item by means of a Boolean 0 – 1 judgement. The judgments so expressed are inserted in the confusion matrix, and then the Cohen's kappa coefficient k_C is computed. As Figure 5 reports, both results are satisfactory, with Functionalities reaching a $k_C = 0.81$, and Quality Criteria with $k_C = 0.86$. Therefore, the PSD structure has been frozen, as it is reported in Figure 6 (B- and C-Matrix) and in Figure 7 (D-Matrix). We note that Figure 6 is a simple adaptation of Figure 2, so we believe no further explanation is necessary; we only make use of colors to stress the connection between project pillars and (Aggregated) items, and gray cells mean that elements in the columns are not related to elements in the rows. Figure 7 reports the D-Matrix, whose yellow cells must be filled in with weights w_{ij} and the a^*_{ij} values, where: $i = 8$ and $j = 19$ for the Technical Pillar, $i = 4$ and $j = 5$ for the Organizational Pillar, and $i = 3$ and $j = 11$ for the Didactical Pillar. This is because a single Functionality can be related to more than just one Aggregated Functionality, and the same applies to Quality Criteria.

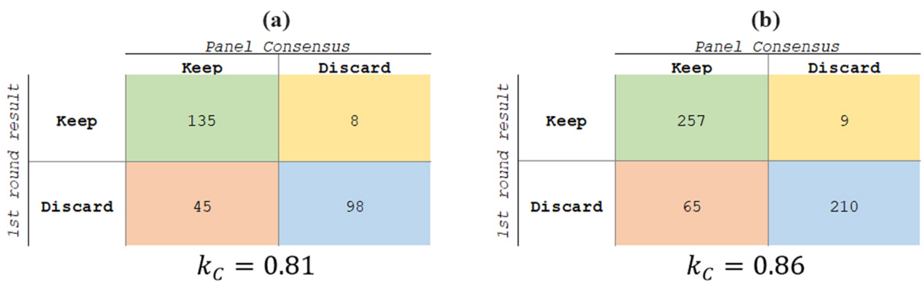


Fig. 5. Kappa coefficient for (a) functionality and (b) quality criteria selections

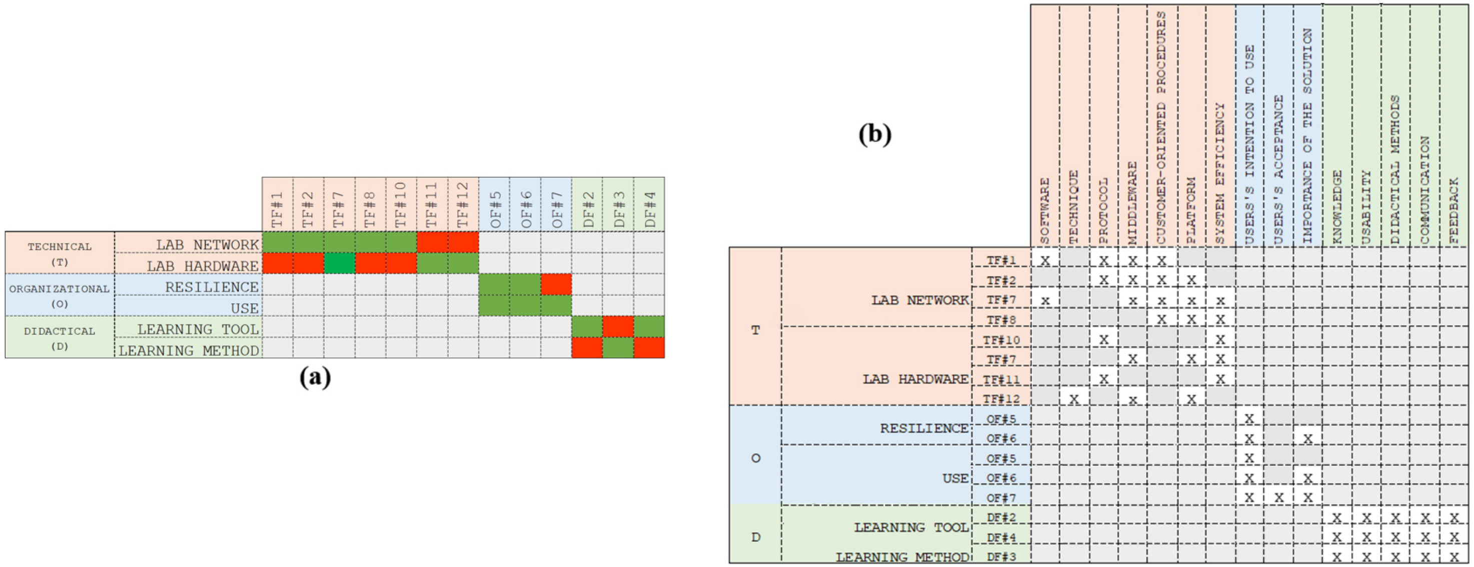


Fig. 6. (a) B-matrix of PSD: aggregated functionalities on the rows and functionalities on the columns; (b) C-matrix of PSD: transpositions of functionalities on the rows, and aggregated quality criteria on the columns

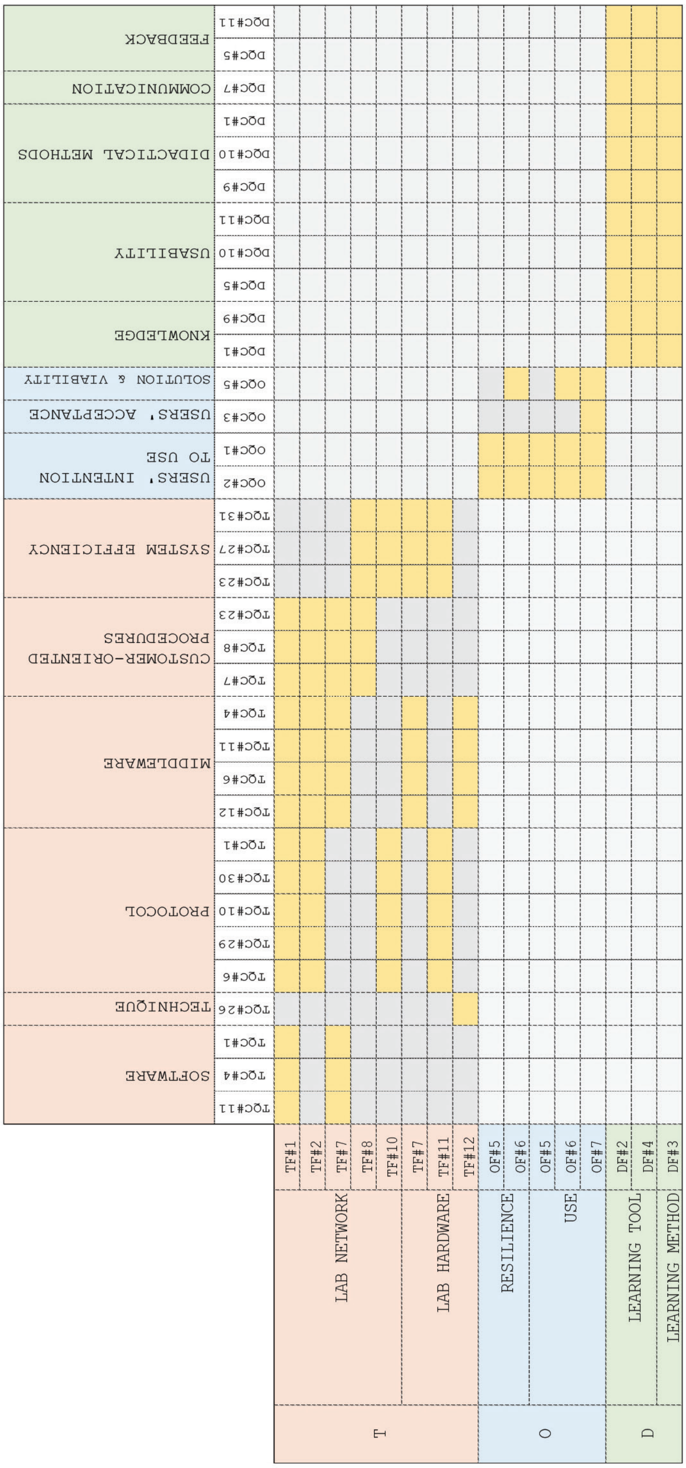


Fig. 7. D-matrix of PSD

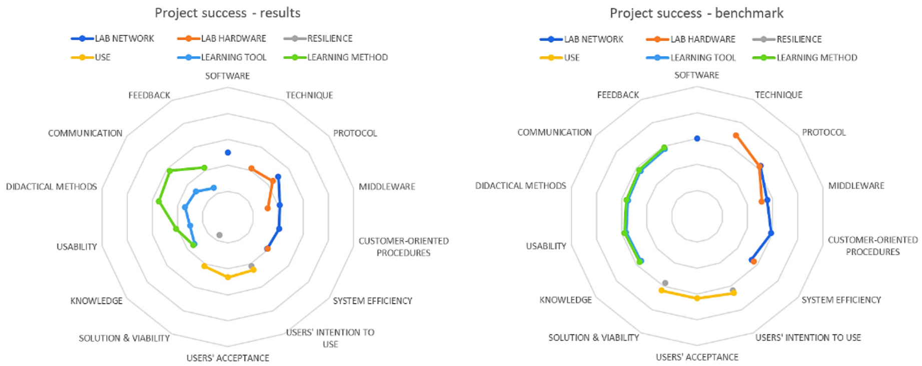
Computations of weights w_{ij} and results achieved by the project. For computing the weights w_{ij} level of importance rated during the selection of items are combined by multiplications as $w_{ij} = \mu_i * \mu_j$. For simplicity, it is arranged a matrix of weights, by putting Functionalities in the rows and Quality Criteria in the columns in the same order in which they are listed in the PSD (see Figure 7).

The further step involves a panel of experts for evaluating the level of achievement of the project, pillar by pillar. We decided to involve a panel of experts that differs from the one who supported the development of the PSD structure. This second panel must fulfil two requirements: (i) experts must be well aware of the project scope, as well as of the project deliverables, to evaluate them in terms of Functionalities and Quality Criteria; (ii) experts must be independent, with their judgements, if not their identity, undisclosed to the project team. Hence, a panel of 15 experts was involved, with competences and skills related to the project pillars. We note that the project steering committee decision of having 5 experts for each project pillar is not mandatory, and we transfer the decision to project evaluators. Each expert evaluated the ratings a_{ij} in a low-medium-high scale (1-3-5), and the level of achievement for the Functionality *i* with respect to the Quality Criterion *j* was then computed from those ratings. We note that, in our specific case, a truncated-mean value of the experts' ratings was used. This decision depends on the fact that the combination of the low number of ratings per field of expertise with the values of the scale led to outlying values in most of the cases. By applying a truncated mean to the 20th percentile, we mitigated the influence of outliers. We note, however, that different sizes of experts' panels, and different scales of judgement could lead to different decisions. Also, the result achieved by the project with respect to the *ith* Functionality and the *jth* Quality Criterion is indicated by $p_{ij} = w_{ij} * a^*_{ij}$. Table A7 (Appendix) provides an example of the results achieved by the Organizational Pillar and the relative benchmark.

Visualization of project success. The global result of the project and its benchmark is reported in Figure 8, which visualizes on a radar chart both the project success indicator $\boldsymbol{S_{ht}}$ and the scope benchmark S_{ht}^{bench} for all Aggregated Functionalities and Aggregated Quality Criteria ($H_f = 6$ and $T_q = 14$, see also Tables A8 and A9 in the Appendix).

For instance, from the technical point of view, the project lags behind with respect to the Aggregated Quality Criterion TECHNIQUE, detailed by the Quality Criterion 'System Reliability'. This result is consistent with the fact that the evaluated digitalization project is releasing 'premature' deliverables, meaning that the digitized labs are still prototypes demonstrated in operational environments, and as such at a Technology Readiness Level below TRL9. Also, another significant item that affects the technical implementation is the MIDDLEWARE, where the project underperforms in terms of the Quality Criterion 'Single Sign-On support', which has not been properly implemented at the moment. We note that this Quality Criterion affects several Functionalities that expected the Single-Sign On implementation.

From the didactical point of view, we note that the project reached interesting results for both Aggregated Functionalities, namely LEARNING METHOD and LEARNING TOOLS, with respect to the Aggregated Quality Criteria DIDACTICAL METHODS and COMMUNICATIONS. On the other hand, other Aggregated Quality Criteria, such as USABILITY, KNOWLEDGE, and FEEDBACK, show room for improvements. With respect to Functionalities and Quality Criteria, we can link these results to the users' experience, namely the possibility to access and use the labs and the didactical contents, as well as the users' satisfaction and the capability of the material to foster motivation and sustain behaviors towards a didactical goal.



(a) (b)

Fig. 8. (a) Result of the open DigiLab4U project, and (b) its benchmark

Moreover, the above-mentioned limits from the didactical point of view are strictly connected to the gaps highlighted in the organizational pillar. The project, in fact, shows non-negligible limits from the organizational point of view. Indeed, the same gaps in terms of motivation to use labs and lab contents that were previously discussed for the didactical pillar do also impact the Functionalities of usability and accessibility of labs, resource availability, and the users' acceptance. Also, other gaps affecting the organizational pillar affect the released system, and its sustainability. The project, in fact, aimed at delivering an economically sustainable lab network. This condition, however, could only be verified after project closure, when the project deliverables are beyond the Go-Live phase (not yet reached), and an enlarged lab network is up and running. Thus, we must note that, although some results are not (yet) encouraging, they are still room for improvement in the last project months, and a timely evaluation can be useful to precisely understand the room for improvement and further development of new releases (e.g., deliverables 2.0), a common approach in digitalization projects.

5 Conclusions

The present paper approaches the problem of evaluating digitalization projects, with a practical application to the evaluation of a digitalization project of laboratories and the development of a digitized lab network. The main novelty of this study is that it introduces a method and a set of indicators for evaluating the success of digitalization projects in terms of how these projects meet their scope, thus answering to the dual need of (i) evaluating the project towards the end of its execution phase, and possibly before the release of all work packages, and (ii) considering both the external and internal dimensions of a project. A project management tool is provided, which we labelled as the Project Successful Deployment (PSD), for evaluating and discussing the results of digitalization initiatives. The design of the tool makes the PSD suitable for any kind of project, although it has been devised and applied in this study to a digitalization project of laboratories. As an example, the PSD could be suitable for assessing digitalization projects in the manufacturing environment, where enabling remote-access, Cloud Manufacturing, and Manufacturing-as-a-Service platforms are at the cutting edge of the fourth industrial revolution.

Our PSD model has been fully tested in a practical case study, namely the Open DigiLab4U project. The case study allowed to evaluate the project, and also to identify strengths and weaknesses of the PSD model. First, the results provided by the PSD are robust, as the statistic test of the model prove, and consistent with the actual status of the project deliverables. For instance, the application of the PSD model identified gaps in the system reliability, as well as potential for further development in terms of feedback, usability, and knowledge, to report some aggregated quality criteria that could be improved. Indeed, the project is still in its execution phase as we write, and there is still room for improvements. On the other hand, digitalization projects are usually affected by the fact that their first release often delivers a work-in-progress product or service, and the evaluation of these results can be biased. This fact is particularly true if we consider the three pillars of the DigiLab4U project: the best results seem to be those of the Technical Pillar, whose deliverable are 'material' and 'quantifiable'. Despite this concern, a strong point of the PSD is that of providing the project steering board and its stakeholders with a timely project evaluation, and with the identification of gaps that could be filled in future project phases, as well as in future releases of the deliverables. In our opinion, this fact has been showed by the analysis of our case study, especially in the evaluation of the results of the organizational and didactical pillars. Hence, we believe that the PSD model could be useful for evaluating digitalization project, and to assess achieved results and areas for future improvement that could be tackled.

Eventually, we note that the PSD model answers to the research questions we posed in the introduction section, by (i) providing a framework for evaluating the success of digitalization projects, in terms of the degree to which Functionalities delivered by the project meet the selected Quality Criteria, and (ii) specifying the details that must be considered to use the model for project evaluation. We note, however, that our model has been conceived for a timely use towards the end of the project execution phase. As such, the evaluation that can be done can hardly provide robust and lasting results, as the only way to assess the value of project deliverables over time is to evaluate their quality, as it is perceived by project users, and possibly in a quantitative way. This, of course, is one of the limits of the model we propose, since some project results can only be understood after the deliverables reach a kind of 'steady state', and therefore the evaluation provided by the PSD could be biased. Nonetheless, we stress the fact that the PSD could be used after a reasonable period of operational time of project deliverables, either as good or services available on the market, or as processes of some digitalized manufacturing process; this can be achieved on the condition of having more information on the deliverables life cycle, as well as some later information on the project results. Finally, two more limits of our model can be listed as follows. First, although we successfully applied the PSD to a digitalization project, we note that much more evidence should be collected, to properly address the effectiveness and the efficiency of our model. Also, as we stated at the beginning of our research, the PSD model emphasizes the quality of a project in delivering its scope. As such, a combination with the evaluation of project time and cost performances could provide a truly comprehensive method of project evaluation.

6 Acknowledgement

This research project was funded by the German Federal Ministry of Education and Research BMBF, grant number 16DHB2116 (Digilab4U – https://digilab4u.com/).

7 References

[1] H. Guo, Z. Yang, R. Huang, and A. Guo, "The digitalization and public crisis responses of small and medium enterprises: Implications from a COVID-19 survey," *Front. Bus. Res. China*, vol. 14, no. 1, p. 19, Dec. 2020. https://doi.org/10.1186/s11782-020-00087-1

[2] Y. Shang, J. Dong, L. Tian, F. He, and J. Tu, "An improved numerical model for epidemic transmission and infection risks assessment in indoor environment," *J. Aerosol Sci.*, vol. 162, p. 105943, May. 2022. https://doi.org/10.1016/j.jaerosci.2021.105943

[3] R. Heradio, L. De La Torre, D. Galan, F. J. Cabrerizo, E. Herrera-Viedma, and S. Dormido, "Virtual and remote labs in education: A bibliometric analysis," *Comput. Educ.*, vol. 98, pp. 14–38, 2016. https://doi.org/10.1016/j.compedu.2016.03.010

[4] R. Heradio, L. de la Torre, and S. Dormido, "Virtual and remote labs in control education: A survey," *Annu. Rev. Control*, vol. 42, pp. 1–10, 2016. https://doi.org/10.1016/j.arcontrol.2016.08.001

[5] G. Esposito, D. Mezzogori, D. Reverberi, G. Romagnoli, M. Ustenko, and F. Zammori, "Non-traditional labs and lab network initiatives: A review," *Int. J. Online Biomed. Eng.*, vol. 17, no. 05, p. 4, May. 2021. https://doi.org/10.3991/ijoe.v17i05.20991

[6] G. Romagnoli, G. Esposito, A. Rizzi, F. Zammori, M. Bertolini, and D. Uckelmann, "Lab networks in engineering education: A proposed structure for organizing information," *Int. J. Online Biomed. Eng.*, vol. 16, no. 05, pp. 41–70, May. 2020. https://doi.org/10.3991/ijoe.v16i05.11891

[7] M. Zappatore, A. Longo, and M. A. Bochicchio, "The bibliographic reference collection GRC2014 for the online laboratory research community," *Proc. 2015 12th Int. Conf. Remote Eng. Virtual Instrumentation, REV 2015*, pp. 24–31, 2015. https://doi.org/10.1109/REV.2015.7087258

[8] E. Bottani, D. Reverberi, G. Romagnoli, M. Ustenko, and A. Volpi, "Batch experiment: A fruitful way of combining hands-on laboratory and e-learning," *Lect. Notes Networks Syst.*, vol. 298, pp. 244–255, 2022. https://doi.org/10.1007/978-3-030-82529-4_24

[9] N. Höhner, A. Pfeiffer, D. Reverberi, M. O. Mints, and J. Rodewald, "Connecting spatially separated laboratory environments by combining virtual and augmented reality technology," 2021. https://doi.org/10.1007/978-3-031-17091-1_51

[10] M. M. Gobble, "Digitalization, digitization, and innovation," *Res. Manag.*, vol. 61, no. 4, pp. 56–59, Jul. 2018. https://doi.org/10.1080/08956308.2018.1471280
[11] K. Achuthan, and S. S. Murali, "A comparative study of educational laboratories from cost & learning effectiveness perspective," in *Software Engineering in Intelligent Systems,* Springer, pp. 143–153, 2015. https://doi.org/10.1007/978-3-319-18473-9_15
[12] Y. Fan, A. Evangelista, and V. Indumathi, "Evaluation of remote or virtual laboratories in e-learning engineering courses," in *2021 IEEE Global Engineering Education Conference (EDUCON)*, Apr. 2021, pp. 136–143, https://doi.org/10.1109/EDUCON46332.2021.9454067
[13] X. Chen, Y. Zhang, L. Kehinde, and D. Olowokere, "Developing virtual and remote undergraduate laboratory for engineering technology," in *2010 Annual Conference & Exposition*, pp. 15–390, 2010.
[14] Nature research custom media and CIDEG, "Expanding innovation through digitalization," *Nature*, 2021.
[15] J. Butt, "A conceptual framework to support digital transformation in manufacturing using an integrated business process management approach," *Designs*, vol. 4, no. 3, p. 17, 2020. https://doi.org/10.3390/designs4030017
[16] G. Dutta, R. Kumar, R. Sindhwani, and R. K. Singh, "Digitalization priorities of quality control processes for SMEs: A conceptual study in perspective of Industry 4.0 adoption," *J. Intell. Manuf.*, vol. 32, no. 6, pp. 1679–1698, 2021. https://doi.org/10.1007/s10845-021-01783-2
[17] S. A. Khan, I. Naim, S. Kusi-Sarpong, H. Gupta, and A. R. Idrisi, "A knowledge-based experts' system for evaluation of digital supply chain readiness," *Knowledge-Based Syst.*, vol. 228, p. 107262, Sep. 2021. https://doi.org/10.1016/j.knosys.2021.107262
[18] S. Thun, O. Bakås, and T. C. B. Storholmen, "Development and implementation processes of digitalization in engineer-to-order manufacturing: Enablers and barriers," *AI Soc.*, 2021. https://doi.org/10.1007/s00146-021-01174-4
[19] J. R. Turner, and R. Müller, "On the nature of the project as a temporary organization," *Int. J. Proj. Manag.*, vol. 21, no. 1, pp. 1–8, 2003. https://doi.org/10.1016/S0263-7863(02)00020-0
[20] Project Management Institute – PMI, *PMI lexicon of project management terms*, 2nd ed. Newtown Square, PA, 2013.
[21] O. Zwikael, and J. Meredith, "Evaluating the success of a project and the performance of its leaders," *IEEE Trans. Eng. Manag.*, 2019.
[22] W. Zhan, W. Pan, and L. Chen, "Construction project productivity evaluation framework with expanded system boundaries," *Eng. Constr. Archit. Manag.*, 2020. https://doi.org/10.1108/ECAM-12-2019-0691
[23] D. Veido, B. Misnevs, and A. Plotkin, "The method of agile projects success evaluation using machine learning," in *International Conference on Reliability and Statistics in Transportation and Communication*, pp. 478–487, 2019. https://doi.org/10.1007/978-3-030-44610-9_47
[24] U. Pinter, and I. Pšunder, "Evaluating construction project success with use of the M-TOPSIS method," *J. Civ. Eng. Manag.*, vol. 19, no. 1, pp. 16–23, 2013. https://doi.org/10.3846/13923730.2012.734849
[25] Z. Huang, M. Poli, and H. S. Mithiborwala, "Project strategy: Success themes for strategic projects," in *PICMET'09-2009 Portland International Conference on Management of Engineering & Technology*, pp. 1282–1289, 2009. https://doi.org/10.1109/PICMET.2009.5261992
[26] M.-Y. Cheng, H.-C. Tsai, and E. Sudjono, "Evolutionary fuzzy hybrid neural network for dynamic project success assessment in construction industry," *Autom. Constr.*, vol. 21, pp. 46–51, 2012. https://doi.org/10.1016/j.autcon.2011.05.011

[27] J. O. Romero, and A. A. Aibinu, "Ex post impact evaluation of PPP projects—The project success evaluation pyramid model," in *INNOVATIVE PRODUCTION AND CONSTRUCTION: Transforming Construction Through Emerging Technologies*, World Scientific, pp. 99–116, 2019. https://doi.org/10.1142/9789813272491_0006
[28] A. De Wit, "Measurement of project success," *Int. J. Proj. Manag.*, vol. 6, no. 3, pp. 164–170, 1988. https://doi.org/10.1016/0263-7863(88)90043-9
[29] N. Agarwal, and U. Rathod, "Defining 'success' for software projects: An exploratory revelation," *Int. J. Proj. Manag.*, vol. 24, no. 4, pp. 358–370, 2006. https://doi.org/10.1016/j.ijproman.2005.11.009
[30] H. Yamashina, and T. Kubo, "Manufacturing cost deployment," *Int. J. Prod. Res.*, vol. 40, no. 16, pp. 4077–4091, 2002. https://doi.org/10.1080/00207540210157178
[31] M. Braglia, M. Frosolini, M. Gallo, and L. Marrazzini, "Lean manufacturing tool in engineer-to-order environment: Project cost deployment," *Int. J. Prod. Res.*, vol. 57, no. 6, pp. 1825–1839, 2019. https://doi.org/10.1080/00207543.2018.1508905
[32] C. Forbes, M. Evans, N. Hastings, and B. Peacock, *Statistical distributions*, vol. 4. Wiley Hoboken, 2011. https://doi.org/10.1002/9780470627242
[33] J. Kotrlik, and C. Higgins, "Organizational research: Determining appropriate sample size in survey research appropriate sample size in survey research," *Inf. Technol. Learn. Perform. J.*, vol. 19, no. 1, p. 43, 2001.
[34] S. Ghorbani, S. E. M. Jafari, and F. Sharifian, "Learning to be: Teachers' competences and practical solutions: A step towards sustainable development.," *J. Teach. Educ. Sustain.*, vol. 20, no. 1, pp. 20–45, 2018. https://doi.org/10.2478/jtes-2018-0002
[35] S. Ahn, S. M. Huff, Y. Kim, and D. Kalra, "Quality metrics for detailed clinical models," *Int. J. Med. Inform.*, vol. 82, no. 5, pp. 408–417, 2013. https://doi.org/10.1016/j.ijmedinf.2012.09.006

8 Appendix

Table A1. Technical functionalities and relative aggregated functionalities

ID	Technical Functionalities	Aggregated Functionalities, Group #1	Aggregated Functionalities, Group #2
TF#1	LabMS	LAB NETWORK	
TF#2	LMS	LAB NETWORK	
TF#3	LRS	LAB NETWORK	
TF#4	Booking	LAB NETWORK	
TF#5	Billing and Payment	LAB NETWORK	
TF#6	Data repository	LAB NETWORK	
TF#7	Interfaces and standardization	LAB NETWORK	LAB HARDWARE
TF#8	Visualization	LAB NETWORK	
TF#9	SCM Serious Game	LAB NETWORK	
TF#10	Security	LAB NETWORK	
TF#11	Safety	LAB HARDWARE	
TF#12	Reliability	LAB HARDWARE	
TF#13	Recovery & Versioning	LAB NETWORK	LAB HARDWARE

Table A2. Organizational functionalities and relative aggregated functionalities

ID	Technical Functionalities	Aggregated Functionalities, Group #1	Aggregated Functionalities, Group #2
OF#1	Absorbing threats	RESILIENCE	
OF#2	Administrative efforts	SUSTAINABILITY	USE
OF#3	Financial Sustainability	SUSTAINABILITY	
OF#4	Trust factors	SUSTAINABILITY	USE
OF#5	Usability and Accessibility	USE	RESILIENCE
OF#6	Resource availability	USE	RESILIENCE
OF#7	Users' acceptance	USE	

Table A3. Didactical functionalities and relative aggregated functionalities

ID	Technical Functionalities	Aggregated Functionalities, Group #1	Aggregated Functionalities, Group #2
DF#1	Collaborative Learning	LEARNING TOOL	
DF#2	Learning Analytics	LEARNING TOOL	
DF#3	Collaborative Learning	LEARNING METHOD	
DF#4	Self-Regulated Learning	LEARNING METHOD	
DF#5	Mixed Reality	LEARNING TOOL	
DF#6	Serious Gaming	LEARNING METHOD	

Table A4. Technical quality criteria and relative aggregated quality criteria

ID	Quality Criteria	Directly Referred To:	Aggregated QC, Group #1	Aggregated QC, Group #2
TQC#1	Use of suitable software and protocols (opensource, standard, largely adopted and maintained by the community)	LAB NETWORK (generic)	SOFTWARE	PROTOCOL
TQC#2	Project methodologies (e.g., agile PM)	LAB NETWORK (generic)	TECHNIQUE	
TQC#3	Documentation and graphic documentation – if useful (e.g., datagram, UML, blocks)	LAB NETWORK (generic)	PROTOCOL	TECHNIQUE
TQC#4	Single Sign-On support (or at least a login system based on federated environments)	LAB NETWORK (generic)	SOFTWARE	MIDDLEWARE
TQC#5	Level of specialization (i.e., a goal-oriented platform, both for software stability and in terms of functionalities it provides)	LMS	MIDDLEWARE	PLATFORM
TQC#6	Configuration (according to platform guidelines)	LMS	PROTOCOL	MIDDLEWARE
TQC#7	Customization of the user experience (e.g., addition of innovative and targeted functionalities, such as booking adapted to the context), and learning data as well	LMS, LRS	CUSTOMER-ORIENTED PROCEDURES	
TQC#8	Easy and robust implementation of new scenarios/courses/activities, and new labs or experiments as well	LMS, LabMS	CUSTOMER-ORIENTED PROCEDURES	
TQC#9	Possibility of using various teaching methods (e.g., self-regulated or collaborative learning), and various didactic tools (e.g., H5P, LA)	LMS	CUSTOMER-ORIENTED PROCEDURES	
TQC#10	Remote application architecture	LabMS	PROTOCOL	MIDDLEWARE
TQC#11	Integration	LabMS, LRS, Booking	SOFTWARE	MIDDLEWARE
TQC#12	Ease of deployment (e.g., docker)	LabMS	MIDDLEWARE	
TQC#13	Easy visualization with advanced dashboards for insights and statistics	LRS, Data repository	CUSTOMER-ORIENTED PROCEDURES	PLATFORM
TQC#14	Easy export of data (at various levels of aggregation)	LRS	CUSTOMER-ORIENTED PROCEDURES	PLATFORM
TQC#15	Easy booking and booking consultation/overview	Booking	MIDDLEWARE	PLATFORM
TQC#16	Management of booking permissions (e.g., by providing figures with different roles)	Booking	MIDDLEWARE	PLATFORM

(Continued)

Table A4. Technical quality criteria and relative aggregated quality criteria *(Continued)*

ID	Quality Criteria	Directly Referred To:	Aggregated QC, Group #1	Aggregated QC, Group #2
TQC#17	Acceptance of standard certified (i) payments (e.g., credit card, PayPal) and (ii) protocols to ensure the security of payment data	Billing and Payment	PROTOCOL	MIDDLEWARE
TQC#18	Drafting of policies for handling "fees" according to national laws	Billing and Payment	PROTOCOL	PLATFORM
TQC#19	Loadable file types	Data repository	PROTOCOL	CUSTOMER-ORIENTED PROCEDURES
TQC#20	Hierarchical organization	Data repository	CUSTOMER-ORIENTED PROCEDURES	PROTOCOL
TQC#21	Standard framework and procedures	Visualization, SCM Serious Game	PLATFORM	SYSTEM EFFICIENCY
TQC#22	Shared layouts and graphics panels in all platform menus	Visualization	PLATFORM	SYSTEM EFFICIENCY
TQC#23	Responsiveness (i.e., possibility to access to platforms by different devices)	Visualization	CUSTOMER-ORIENTED PROCEDURES	SYSTEM EFFICIENCY
TQC#24	Diary with recovery simulations	SCM Serious Game, Recovery & Versioning	PLATFORM	
TQC#25	Versioning of sources and documents with dedicated tools	SCM Serious Game, Recovery & Versioning, Reliability	PLATFORM	
TQC#26	System Reliability	Security, Reliability	TECHNIQUE	
TQC#27	Down time, Deliveries on times, Recovery time	Security	SYSTEM EFFICIENCY	
TQC#28	Cost to keep automatic	Security	SYSTEM EFFICIENCY	
TQC#29	Privacy	Security	PROTOCOL	
TQC#30	Certificates	Safety	DOCUMENTATION	
TQC#31	Effectiveness (degree to which safety measures counteract the threats)	Safety	SYSTEM EFFICIENCY	
TQC#32	Conformity (dependent on the quality of the implementation)	Safety	SYSTEM EFFICIENCY	

Table A5. Organizational quality criteria and relative aggregated quality criteria

ID	Quality Criteria	Directly Referred To:	Aggregated QC, Group #1	Aggregated QC, Group #2
OQC#1	Usability (by System Usability Scale – SUS metrics)	Usability and Accessibility, User's acceptance, Resource availability	USERS' INTENTION TO USE	
OQC#2	Motivational aspects	Users' acceptance	USERS' INTENTION TO USE	
OQC#3	Trustworthiness (by SCOR metrics)	Trust factors, Users' acceptance	USERS' ACCEPTANCE	
OQC#4	Relevance	Absorbing threats, administrative efforts	IMPORTANCE OF THE SOLUTION	
OQC#5	Sustainability	Users' acceptance, Resource availability, Financial Sustainability	IMPORTANCE OF THE SOLUTION	VIABILITY
OQC#6	Willingness to pay	Financial Sustainability	VIABILITY	

Table A6. Didactical quality criteria and relative aggregated quality criteria

ID	Quality Criteria	Directly Referred To:	Aggregated QC, Group #1	Aggregated QC, Group #2
DQC#1	Technical competence required	Serious Gaming, Learning Analytics, Mixed Reality	KNOWLEDGE	DIDACTICAL METHODS
DQC#2	Lab booking effectiveness	Learning Analytics, Collaborative Learning, Self-Regulated Learning	AVAILABILITY	KNOWLEDGE
DQC#3	Preparation time for using labs	Collaborative Learning, Self-Regulated Learning	FLEXIBILITY	
DQC#4	Adaption to pedagogical concerns	Learning Analytics, Collaborative Learning, Self-Regulated Learning	USABILITY	FEEDBACK
DQC#5	Availability of didactical material	Collaborative Learning, Self-Regulated Learning	USABILITY	FEEDBACK
DQC#6	Organization of didactical material	Collaborative Learning, Self-Regulated Learning, Learning Analytics	DIDACTICAL METHODS	COMMUNICATION
DQC#7	Documentation for using labs and lab material	Collaborative Learning, Self-Regulated Learning	COMMUNICATION	
DQC#8	Improvement of material and didactical tools	Learning Analytics, Open Badges	FEEDBACK	
DQC#9	Learning outcome	Learning Analytics, Open Badges	DIDACTICAL METHODS	KNOWLEDGE
DQC#10	Learning motivation	Learning Analytics	USABILITY	DIDACTICAL METHODS
DQC#11	Users' satisfaction	Learning Analytics	USABILITY	FEEDBACK

Table A7. Results achieved by the organizational pillar and relative benchmark values (lower line of each row, reported with gray-colored font)

		USERS' INTENTION TO USE		USERS' ACCEPTANCE	SOLUTION & VIABILITY
Organizational WBS		**Motivational Aspects**	**Usability (by System Usability Scale – SUS Metrics)**	**Trustworthiness (by SCOR Metrics)**	**Sustainability**
Resilience	***Usability and Accessibility***	2.23	2.32		
		3.05	3.87		
	Resource availability	1.87	1.94		0.76
		2.55	3.23		2.86
Use	***Usability and Accessibility***	2.23	2.84		
		3.05	3.87		
	Resource availability	1.70	1.94		2.09
		2.55	3.23		2.86
	Users' acceptance	2.50	2.38	2.32	2.10
		3.13	3.97	3.17	3.51

Table A8. Global result of the project: project success indicator S_{ht}

Global Result of the Project – Fulfilled Scope	Software	Technique	Protocol	Middleware	Customer-Oriented Procedures	System Efficiency	Users' Intention To Use	Users' Acceptance	Solution & Viability	Knowledge	Usability	Didactical Methods	Communication	Feedback
LAB Network	2.50		2.50	2.07	2.04	1.95								
LAB Hardware		2.09	2.23	1.58		1.96								
Resilience							2.09		0.76					
Use							2.26	2.32	2.10					
Learning Tool										1.66	1.50	1.70	1.59	1.25
Learning Method										1.72	2.05	2.74	2.87	2.12

Table A9. Global result of the project: scope benchmark S_{ht}^{bench}

Global Result of the Project – Desired Scope	Software	Technique	Protocol	Middleware	Customer-Oriented Procedures	System Efficiency	Users' Intention To Use	Users' Acceptance	Solution & Viability	Knowledge	Usability	Didactical Methods	Communication	Feedback
LAB Network	3.01		3.14	2.79	2.95	2.69								
LAB Hardware		3.48	3.09	2.57		2.81								
Resilience							3.17		2.86					
Use							3.30	3.17	3.18					
Learning Tool										2.78	2.82	2.75	2.81	2.88
Learning Method										2.84	2.88	2.81	2.87	2.94

9 Authors

Giovanni Romagnoli is a lecturer in Operations and Project Management at the Department of Engineering and Architecture of the University of Parma. Since December 2019, he is the local scientific coordinator of the project DigiLab4U (http://digilab4u.com/). His research interests include virtual and remote lab network initiatives, RFID and supply chain management, production planning & control systems. He has published more than 50 contributions on International Journals or Conference proceedings, some of which have been awarded with prizes. His ORCID iD is: https://orcid.org/0000-0002-9891-0314 (email: giovanni.romagnoli@unipr.it).

Giovanni Esposito obtained a PhD in Industrial Engineering at the University of Parma (Italy) in 2021. His research interests comprehend RFID application for supply chain management and supply chain modelling towards Industry 4.0. He also visited the Stuttgart University of Applied Sciences (Germany). He acted as a local manager for the University of Parma on the DigiLab4U project. His ORCID iD is https://orcid.org/0000-0001-5150-0855.

Davide Reverberi was born in 1992 in Parma (Italy). He completed his master's degree in Management Engineering with a focus on production methodologies at the University of Parma. He is now a research fellow at the University of Parma at the Department of Engineering and Architecture. Since May 2019, he has been involved in the DigiLab4U project with a focus on remote labs and lab-based learning. During this period, he has attended several workshops, both in Italy and abroad, on technical and didactic topics for labs. He is currently co-author of eight international scientific publications, two of which have been accepted in international journals. He is local administrative manager for the project DigiLab4U. His ORCID iD is: https://orcid.org/0000-0001-6768-3932

Article submitted 2022-09-02. Resubmitted 2022-10-09. Final acceptance 2022-10-15. Final version published as submitted by the authors.